Preferential market access

SUMMARY

This paper suggests an integrated approach to study selection into and consequences of five modes of preferential economic integration agreements (PEIAs): goods trade agreements (GTAs), services trade agreements (STAs), double taxation treaties (DTTs), bilateral investment treaties (BITs), and currency unions as well as currency pegs (CUAs). A detailed descriptive analysis reveals typical integration patterns, with DTTs and BITs often being first steps towards deeper integration. We consider the effects of PEIAs on bilateral goods trade, services trade, and FDI and provide conclusive evidence that single and combined PEIAs are associated with positive effects not only on single outcome but typically on all outcomes. Investment liberalization through DTTs and BITs seems to be particularly beneficial since concluding them alone or in any combination with other agreements encourages goods trade even more than the liberalization of goods trade per se.

— Peter Egger and Georg Wamser

Economic Policy January 2013 Printed in Great Britain

Multiple faces of preferential market access: their causes and consequences

Peter Egger and Georg Wamser

ETH Zürich, WIFO, CEPR, CESifo, GEP and OUCBT; University of Tübingen and CESifo

1. INTRODUCTION

World War II reset the world economy almost to zero economic integration relative to how integrated the world was at the beginning of the twentieth century. Many if not most of the bilateral and multilateral agreements granting preferential access to foreign markets – through trade in goods and services but also along dimensions beyond such transactions – were abandoned in its course. Accordingly, multilateral political and economic cooperation through the General Agreement on Trade and Tariffs (GATT) in 1947 as well as preferential integration through the formation of the European Community in 1958 (Treaties of Rome) and the European Free Trade Association (EFTA) in 1960 happened almost in a vacuum of integration of that kind.

We gratefully acknowledge numerous helpful comments from three anonymous reviewers, the editor in charge, as well as discussants and participants at the 55th Economic Policy Panel in Copenhagen. We are indebted to members of the WTO for sharing their classification scheme of goods trade and service trade agreements with us. In that regard, special thanks go to Rudolf Adlung, Alen Mulabdic, Roberta Piermartini, Nadia Rocha and Pierre Sauvé, for providing us with data. An earlier draft of the paper has benefited by the invaluable comments of Rudolf Adlung, Richard Baldwin, Petros Mavroidis and André Sapir.

The Managing Editor in charge of this paper was Philip Lane.

Economic Policy January 2013 pp. 143–187 Printed in Great Britain

While the political obstacles to those agreements were big at the time, so were the economic merits, given the level of disintegration brought about by the War. At that point of history, had data on bilateral goods trade been as abundant as nowadays, the trade gains from preferential liberalization would have been almost directly attributable to the inception of those agreements.

In terms of preferential as well as non-preferential economic integration, the world has fundamentally changed since. For instance, tariffs have been substantially reduced within GATT (now the members of the World Trade Organization, WTO) through seven completed rounds of negotiations mainly about tariffs. Since the completion of the Uruguay round in 1986, most countries apply single-digit average most-favoured nation tariff rates negotiated under GATT not only to members of the WTO but even to non-members. Moreover, since the early 1990s the number of preferential trade agreements (both under and outside of the auspices of WTO) skyrocketed. As of November 2011, 313 regional (goods and services) preferential trade agreements notified to the WTO were in force, with many more being negotiated. While economic research often focuses on and emphasizes the role of goods trade liberalization, neither preferential nor multilateral liberalization of goods trade happened in isolation. The foundation of the General Agreement on Trade in Services (GATS) in 1995 paved the way for liberalization of (mostly) trade and commercial presence in services which saw major steps forward not only in the documentation of the degree of liberalization of service transactions across countries and sectors in 1996 but also in the conclusion of preferential agreements about services trade. Also the latter happened mostly since the 1990s.[1]

While goods trade agreements (GTAs) and services trade agreements (STAs) were arguably important for either type of trade, growth of trade has long been outpaced by the growth of multinational enterprise (MNE) activity over the last three decades (see Markusen, 1995). Economic theory suggests that multinational activity induces trade in (headquarters) services which either complements (through vertical multinational organization) or substitutes (through horizontal multinational organization) trade in goods (see Helpman, 1984; Markusen, 1984, 2002; Markusen and Venables, 1998, 2000; Ekholm *et al.*, 2007). Hence, GTAs are a potential driver of goods trade, and STAs potentially affect services trade, but they may also interact and affect outcome through the presence of vertical, horizontal, or more complex forms of MNEs. Two other modes of preferential liberalization directly address and affect the environment of multinationals in their foreign activity: bilateral investment treaties (BITs) specifically regulate procedures in the case of expropriation and generally influence the risk of bilateral investments; and double taxation treaties (DTTs), which aim at avoiding the double taxation of profits earned abroad (apart from ensuring informa-

[1] Of the 313 aforementioned preferential trade agreements notified to the WTO and in force by November 2011, 122 were pure goods trade agreements, 1 was a pure services trade agreement, and 88 were agreements covering goods and services trade.

tion exchange between taxing authorities). By virtue of the interdependence of trade and MNE activity, these agreements may display direct or indirect effects on trade flows in goods or services. Beyond those four pillars of preferential liberalization, currency unions in a broad sense (covering currency unions as well as currency pegs; CUAs) affect economic transactions of any considered kind (trade in goods and services as well as MNE activity) through a reduction of exchange rate risk in a quasi-preferential way.

Both normative and positive work on the one hand, and both theoretical and empirical work on the other hand has recognized the role of all of these modes of preferential liberalization for long. At the same time, research on these matters of integration has been conducted in an astonishingly disintegrated way. The analysis of (mostly) one single mode of economic integration is symptomatic for the majority of the theoretical as well as empirical work available. While some authors have paid specific attention to the interaction between some pairs of modes, their interest was typically of a theoretical nature and related empirical work is not or only scarcely available. It is common in quantitative work to assume that preferential agreements of the mentioned kind are concluded independently both across treatments and in time so that partial inference can inform us about their marginal impact. Hence, the academic approach in economics and political science to the causes and consequences of Preferential Economic Integration Agreements (PEIAs) is mainly unimodal in two ways: it focuses on one type of agreement at a time and, when assessing the consequences, considers only one outcome. The latter is potentially harmful, because we may misattribute effects on a given outcome to certain agreements or underestimate effects of agreements at large.

Figure 1 considers preferential economic integration in goods trade agreements (GTAs), services trade agreements (STAs), double taxation treaties (DTTs), bilateral investment treaties (BITs), and currency union agreements, including currency pegs (CUAs) among 210 economies[2] and 43,890 dyads (country-pairs) over the time period 1960 to 2005.[3] In particular, the figure shows how the number of PEIAs concluded developed over time for each type of PEIA. It suggests that double taxation treaties outnumber trade agreements of any kind by far. The figure shows an ever-increasing number of all types of agreements except currency unions and pegs over the sample period 1960 to 2005. While currency unions and pegs were the most common form of

[2] The list of countries includes self-governing economies (e.g. the Channel Islands) as separate entries.

[3] The definition of PEIA variables is as follows. GTAs are customs unions, free trade areas, or what the WTO refers to as preferential trade agreements covering goods trade issues; the latter being more shallow than the former ones. The source of information underlying the binary GTA variable is taken from the WTO and the dataset collected by Egger and Larch (2008). STAs are agreements notified to the WTO that accord with Article 92 of the GATS (most of them overlapping with GTAs). Information on DTTs and BITs was collected from the United Nation Conference on Trade and Development's (UNCTAD's) online database on DTTs and BITs. The binary CUA variable was constructed on the basis of the data on *de facto* exchange rate arrangements collected by Reinhart and Rogoff (2004) and updated based on information from the International Monetary Fund (IMF).

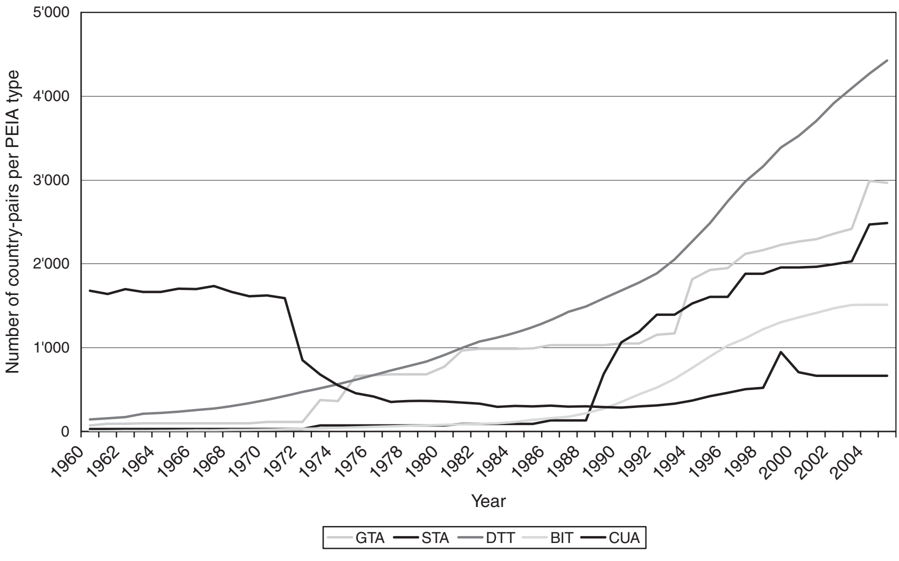

Figure 1. Number of PEIAs 1960–2005

integration until 1974, this form of integration became less important during the last decades.[4]

Figure 1 also demonstrates that the lion's share of the considered agreements came into force in the 1990s and thereafter. As such, this does not cast doubt on the unimodal approach about PEIAs taken by economists. However, Table 1 illustrates that the five modes of PEIAs are by no means mutually exclusive.

Table 1 is organized as follows. The rows condition on an observation (i.e. a country-pair-year unit) to have a specific agreement and the cells infer how big of a percentage of the observations had agreements as listed in the columns. For instance, the cell in the lower right corner, '*None–None*' indicates that about 92% of all observations did not have any of the considered agreements in place. Hence, on average across all years, 8% of the observations had some type of agreement. The cell *STA-GTA* in the upper left corner suggests that about 0.6% of all country-pairs operated under both a services trade agreement and a goods trade agreement in the average year. Double taxation treaties come relatively frequently in isolation (2.2%), together with goods trade agreements (0.5%), or together with bilateral investment treaties (0.6%). Bilateral investment treaties are unlikely to overlap with goods trade agreements and services trade agreements (either of them about 0.1%). The reason for the latter may be that some goods trade agreements and services trade agreements

[4] Note that over the last years (2003–2005) in the sample, about 10% of the 43,890 country-pairs operated under double taxation treaties; about a third as many pairs had bilateral investment treaties (compared with double taxation treaties); only about 5–6% of the pairs had goods or services trade agreements and less than 2% of all pairs had a currency union or a peg.

Table 1. A snapshot of PEIA overlap

Countries with/and	GTA	STA	DTT	BIT	CUA	None
GTA	.	.	.	.	.	
STA	0.006	.	.	.	.	
DTT	0.005	0.004	.	.	.	
BIT	0.001	0.001	0.006	.	.	
CUA	0.002	0.001	0.002	0.0003	.	
None	0.015	0.009	0.022	0.003	0.015	0.921

Notes: 2,018,940 observations, 43,890 country-pairs, 1960–2005. Values reflect average frequencies as fractions of all observations *with* PEIA (row) *and* PEIA (column). GTAs are Goods Trade Agreements; STAs are Services Trade Agreements; DTTs are Double Taxation Treaties; BITs are Bilateral Investment Treaties; and CUAs are Currency Union Agreements which include unions with one currency as well as currency pegs.

reduce the risk of expropriation of foreign investors similar to bilateral investment treaties so that they substitute for having a bilateral investment treaty (e.g. this is the case for the European Union).

This paper aims at assessing the fundamental variables determining selection into the five modes of economic integration in the panel dataset based on 46 years, 210 economies, and 43,890 dyads (country-pairs) underlying Table 1. Selection in five modes of *mutually non-exclusive* PEIAs spans a possible set of $2^5 = 32$ *mutually exclusive* preferential treatment configurations. Of those 32 configurations, only 26 are actually used among the 210 economies between 1960 and 2005. We use multivariate nonlinear probability models to estimate probabilities for the mutually exclusive treatment combinations used by country-pairs. Based on these probabilities and an approach of selection on observables to reduce the self-selection bias of country-pairs into treatments through matching, we assess effects of PEIA configurations on six different outcome variables: the probability of positive *bilateral goods trade*, the level of bilateral goods trade, the probability of positive *bilateral services trade*, the level of bilateral services trade, the probability of positive *bilateral FDI*, and the level of bilateral FDI.[5] We first estimate long-run responses of outcome to selection into different treatments as of the year 2005. Considering causal effects of treatment on outcomes in that year corresponds to a long-run analysis, since we do not require outcome to respond to treatment within a certain (short) time span, but the response may have taken years if not decades. In a second step, we isolate typical patterns of integration paths found in the data. For instance, there is to date only a small number of country-pairs which

[5] The effects we estimate should be thought of as *direct effects*, not accounting for general equilibrium repercussions. Appendix B indicates how those direct treatment effects relate to *total effects* in general equilibrium, and it illustrates why, on average, the two are quite similar. The direct effects of PEIAs estimated here consist of two components, an *immediate effect* and a *mediated effect* on outcome. For instance, GTAs on the one hand reduce tariffs (the immediate effect) but on the other hand often induce institutional changes which go beyond mere trade policy but may influence trade (the mediated effect). This has to do with what Richard Baldwin (2011) calls 21st-century regionalism where trade preferentialism is used to determine rules (about competition, the environment, services, labour standards, etc.) which may have indirect effects on trade.

liberalize services trade without making goods trade provisions. In general, there is a tendency to liberalize goods trade first and services trade in the aftermath. It is less clear as to when countries liberalize foreign direct investment through bilateral investment treaties and double taxation treaties relative to the implementation of goods trade agreements and services trade agreements. We may contrast archetypal integration paths in terms of their effects on outcome with alternative integration paths. This may inform us about *preferable* integration paths on average, at least among a subset of alternatives. We may ask the question how PEIA switching affects outcome relative to non-switching within reasonable time windows by conditioning on the same PEIA treatment configurations for switching and non-switching country-pairs. The latter approach delivers insights into the magnitude of short-run versus medium-run responses to the inception of PEIAs and the speed of adjustment of outcomes towards a new counterfactual equilibrium.

Compared to the existing literature, this analysis covers a number of novel features including the consideration of multiple PEIA treatments, of multiple outcomes, of the extensive versus the intensive country margin of these outcomes, and dynamic adjustments of those margins. In particular, the paper has two goals. First, to indicate which initial integration strategy average country-pairs should adopt when having maximization of a specific outcome in mind. Second, to reveal which integration path country-pairs eventually should pursue with such an objective in mind. The latter also relates to both the sequence and the number of types of PEIAs country-pairs should implement.

In the next section, we proceed with an eclectic overview of the literatures on individual types of PEIAs as considered here. This helps us formulating reduced-form models which portray the net gains for countries from selecting into one of the five PEIAs and combinations thereof. Section 3 dissects data on PEIA treatments. Section 4 alludes to the econometric approach adopted in this paper to reduce the potential self-selection bias associated with PEIA treatments on cross-border aggregate economic activity. Section 5 describes the data sources not only for PEIAs and outcomes but also for their determinants. Section 6 presents estimation results of the multivariate selection models along with estimates of the impact of treatments on outcomes. The last section concludes with a summary of the main findings.

2. AN ECLECTIC LITERATURE REVIEW

For reasons of brevity, let us focus on PEIAs which intend to raise welfare through a stimulus of trade or foreign direct investment. From a general viewpoint, theoretical models about the (local or global) welfare effects of PEIAs suggest that they are larger if the effects of PEIAs on targeted outcome (trade with trade-related agreements and foreign direct investment with investment-related ones) are larger and undesired side effects (diversion of trade or investment) are smaller. Under which economic circumstances this is the case depends on the fundamental reasons for trade or investment: differences in relative factor endowments in Heckscher–Ohlin-type models; differences in technology in Ricardian models; total and relative country size, trade and investment costs, as well as price mark-ups over marginal costs in new trade theory models; market structure and elasticity of demand, market size, and trade and investment costs in oligopoly and other variable-price-elasticity-of-demand models of trade and investment.[6] In all such models, the feasibility of agreements depends on whether compensating transfers among the parties negotiating about an agreement are possible or not.

Most of the empirical literature on PEIAs builds on static new trade theory models of national or multinational firm activity (such as trade or foreign direct investment, FDI). The reason for this is that new trade theory models suggest gravity-type models of bilateral gross flows of trade or investment for which there appears to be overwhelming support by the data: the bilateral volume of trade and investment can be explained with great success by employing exporter plus importer country size (in terms of GDP) and exporter-to-importer relative country size along with measures of trade and investment costs (with alternative functional forms of using country size variables). Relative factor endowment differences or technology differences seem much less important in relative terms. Since the theory of PEIAs suggests that the same fundamentals that determine trade and FDI should determine the welfare effects of PEIAs (such as GTAs, STAs, BITs or DTTs), larger, similarly sized countries with higher trade or investment costs before, and smaller trade and investment costs after concluding a PEIA, should be more inclined towards participating than others. While there are further aspects important for the formation of CUAs, relating to price volatility and economic stability, essentially the same determinants motivating GTAs or STAs as well as BITs or DTTs also describe the features of an optimal currency area, so that they apply for tying currencies, too (see Dorn and Egger, 2011). Prototypes of new trade theory based work on the determinants and consequences of trade-related PEIAs which find support of the just-mentioned relationships for goods trade are Baier and Bergstrand (2004, 2007, 2009) and Magee (2003). In essence, these authors find that larger, more similarly sized countries with lower trade costs in the absence of political barriers to trade conclude trade-related PEIAs (mostly GTAs) more likely, and that such agreements have the intended stimulating effect on trade. While, at this point, there is less systematic evidence about the causes and consequences of STAs, it seems that liberalization of services trade is driven by the same factors and in a similar way as goods trade liberalization (see Egger and Lanz, 2008; Francois and Hoekman, 2010).

There is evidence at both the aggregate level of investment (see Egger and Pfaffermayr, 2004) as well as the firm level (see Egger and Merlo, 2012) that BITs encourage

[6] There is another important strand of work which relates to the role of political economy factors for protection. However, we will not touch upon this topic since measurement of the fundamentals for a large cross-section of countries is difficult (see Arcand *et al.*, 2010, for political economy determinants of trade agreements).

FDI. More importantly, in our context, new trade theory based empirical work on the causes of BITs suggests that they, akin to trade agreements, are concluded among partners which should display larger volumes of bilateral FDI in the absence of political or risk-related barriers to MNE activity. For instance, Bergstrand and Egger (2011) find that country size as well as trade and investment costs are the key drivers of the conclusion of (GTAs and) BITs. As with GTAs only, there is some role to play for relative factor endowment differences but the propensity of concluding a GTA, a BIT, or both, for any country-pair, is largely dominated by economic size of the integrating market and by barriers to trade and investment.

The consequences of DTTs are less clear-cut for an obvious reason: on the one hand, DTTs aim at avoiding discriminatory (e.g. between exporting and multinational firms) double taxation of foreign-earned profits, which, *ceteris paribus*, should stimulate FDI; on the other hand, DTTs aim at greater transparency about domestic and foreign tax bases of national investors and intend to close loopholes in profit taxation and to reduce opportunities of tax fraud of international investors. This means that, from an investor's perspective, the net stimulus on FDI is unclear (see Blonigen and Davies, 2004). But it is still the case that DTTs are concluded more likely among such countries where bilateral FDI should be large according to new trade theory and, *ceteris paribus*, they tend to cause bigger volumes of bilateral FDI on average (see Egger *et al.*, 2006).

Finally, there is broad evidence of a positive effect of tighter currency alignments for both trade (see Rose, 2000; Glick and Rose, 2002; Egger, 2008) and investment (see Goldberg and Kolstad, 1995). Countries tend to self-select into currency unions and pegs systematically, broadly in line with optimum currency area criteria (see Persson, 2001; Barro and Tenreyro, 2007). In new trade theory models, the latter are largely consistent with factors determining greater flows of goods, services and, eventually, factors (such as investment). Thus, on average, we should find CUAs being concluded more likely among large, similarly sized countries with smaller trade (and investment) costs.

This brief overview suggests that new trade theory fundamentals matter for both trade and MNE activity such as FDI and, hence, they determine preferential integration through GTAs, STAs, BITs, DTTs, and even CUAs. Yet, while the same fundamentals appear to be drivers of all such PEIAs, empirical research on their causes and consequences did not strive for an integrated approach in their analysis so far. Implicitly, it has been assumed that such agreements are concluded independently of each other and that they do not influence each other's impact in determining outcome. We provide evidence for a tremendous overlap in the conclusion of different types of PEIAs which suggests that changes in outcome can not be trivially ascribed to one or the other type of agreement. Moreover, not only do bilateral goods trade, services trade, and FDI tend to be largest for the same countries, a significant part of trade of all kinds is even conducted directly by multinational firms. Thus, focusing on effects of GTAs on goods trade alone or of BITs and DTTs on FDI alone will likely miss out important effects on other international economic outcomes. This calls for an integrated approach to consider effects of PEIAs on outcome(s).

3. SOME FACTS ABOUT PEIA MEMBERSHIP

While Figure 1 and Table 1 shed some light on the frequency of PEIAs and the overlap of alternative modes thereof, deeper insights can be gained from a more systematic descriptive analysis. For instance, when taking the yearly numbers on PEIAs presented in Figure 1 relative to the total number of country-pairs per year (43,890), we may obtain the probability of a randomly drawn country-pair to be engaged in one or the other agreement. Table 1 indicates the probability of a randomly drawn pair to combine one type of PEIA with another one. However, this does neither allow conclusions about typical multi-modal or single-modal PEIAs in the data nor does it provide insights into typical integration paths of country-pairs. We shed light on these matters in this section.

Consider the aforementioned five modes of PEIAs (GTAs, STAs, DTTs, BITs, and CUAs) and let us focus on the frequency of such agreements of the 210 countries with the other 209 countries (i.e. of 43,890 country-pairs) over the period 1960–2005.[7] Throughout the paper, we use binary indicators for DTTs, BITs, GTAs, and STAs which are unity whenever a given agreement was signed – in or prior to a respective year (e.g. as notified to the WTO).[8] Altogether, the dataset considered includes 2,018,940 country-pair-year data points. Obviously, since PEIAs of the considered type are symmetric, one could drop half of the observations without any loss of insight. However, we will keep all observations for the ease of data handling when estimating consequences of PEIAs later on.[9] Table 2 presents all $2^5 = 32$ PEIA-combinations countries might principally use, and it accumulates observations across all years in the sample. For better illustration, we sort combinations by their frequency in the data.

[7] While we cover five modes of PEIAs here, one could think of both deepening and broadening even this large set. For instance, differentiating between trade agreements – say, customs unions, free trade areas, and partial scope agreements as notified to the WTO – rather than subsuming all of those agreements under the heading of GTAs may be useful. Alternatively, one could distinguish between deep and shallow GTAs as in Horn *et al.* (2010). We could refer to such approaches as ones of deepening the notion of GTAs. For instance, in an earlier draft of this manuscript, we used deep versus shallow GTAs rather than a single type (results are available from the authors upon request). Moreover, one could pursue a broader approach by covering also other agreements such as ones on competition, environmental standards, or labour standards. Any one of those alternatives would lead to a non-trivial increase in the number of treatments, with increased problems for precise estimation. Hence, we focus on the proposed catalogue of PEIAs which is already much bigger than the ones considered in previous research.

[8] At least for DTTs and BITs, it turns out that investment effects occur prior to the date these treaties come into force, and that these effects are large relative to the entry-into-force effects. With such anticipation effects, the total treatment effect would be downward biased (in labour economics, this is referred to as Ashenfelter's dip) when focusing on entry-into-force effects. Similarly, many trade agreements involve phase-in adjustments of tariffs or institutions prior to the agreement itself coming into effect. Clearly, it is a judgment call which one of the two to focus on, but we felt more comfortable with using PEIA signatures in quantifying effects of PEIAs (rather than ratifications).

[9] One might be worried about a deflation of standard errors in the PEIA selection models, but those could easily be adjusted *ex post*, if desired. In any case, while the treatments considered and the treatment effects estimated are symmetric between pair *ij* and pair *ji* in any given year, the outcomes for those two pairs are not the same.

Table 2. Being in, switching into, and switching out of PEIAs

PEIA treatments							
BIT	STA	DTT	BIT	CUA	Observations (total)	Switchers into	Switchers out of
0	0	0	0	0	1,859,225	2,405	8,536
0	0	1	0	0	44,820	3,165	1,055
0	0	0	0	1	30,041	1,055	2,354
1	0	0	0	0	29,549	2,021	773
0	1	0	0	0	17,961	1,379	163
0	0	1	1	0	9,990	990	273
0	0	0	1	0	5,721	739	488
1	1	0	0	0	4,833	462	161
1	0	1	0	0	4,005	510	98
1	1	1	0	0	3,696	489	122
0	0	1	0	1	1,903	189	169
1	1	1	0	1	1,482	185	13
1	0	0	0	1	1,087	56	58
0	1	1	0	0	888	105	35
1	1	1	1	0	732	217	4
1	0	1	1	0	725	138	34
0	0	1	1	1	484	76	10
1	1	0	0	1	393	46	40
0	1	0	1	0	354	50	12
1	0	0	1	0	327	56	28
0	1	1	1	0	248	33	5
1	1	0	1	0	241	61	19
0	0	0	1	1	158	28	9
1	1	1	1	1	60	4	0
0	1	0	0	1	16	2	2
1	0	1	0	1	1	1	1
0	1	0	1	1	0	0	0
0	1	1	0	1	0	0	0
0	1	1	1	1	0	0	0
1	0	0	1	1	0	0	0
1	0	1	1	1	0	0	0
1	1	0	1	1	0	0	0
					2,018,940	14,462	14,462

Notes: GTAs are Goods Trade Agreements; STAs are Services Trade Agreements; DTTs are Double Taxation Treaties; BITs are Bilateral Investment Treaties; and CUAs are Currency Union Agreements which include unions with one currency as well as currency pegs.

The figures in Table 2 can be interpreted as follows. First, of the 32 possible combinations of PEIAs, only 26 actually appear in the data. Of all the combinations possible, the no-PEIA scenario in the top row is used most frequently (notice that this is largely due to a small number of PEIAs in the early decades of the sample period). When drawing observations randomly from the data, there would be a chance of $100 \times 1{,}859{,}225/2{,}018{,}940 \approx 92\%$ to draw a unit where no facilitated market access of any considered kind is granted to firms in two countries vis-à-vis each other. Notice that only in 60 cases country-pairs in the data use every type of PEIA mode possible. There are numerous reasons for why this is the case. For instance, some GTAs and STAs include investment provisions which render regulations usually formulated in

Table 3. Switching out of/into PEIA treatments

Out/In	00000	00001	00010	00011	00100	00101	00110	00111	01000	01001	01010	01011	01100	01101	01110	01111
00000	0	1,055	736	4	2,995	8	139	0	1,371	0	2	0	1	0	0	0
00001	2,132	0	0	7	14	103	0	0	0	2	0	0	0	0	0	0
00010	0	0	0	17	0	0	393	1	0	0	3	0	0	0	0	0
00011	0	0	3	0	0	0	0	6	0	0	0	0	0	0	0	0
00100	0	0	0	0	0	78	433	6	0	0	0	0	51	0	2	0
00101	0	0	0	0	123	0	2	10	0	0	0	0	0	0	0	0
00110	0	0	0	0	0	0	0	53	0	0	0	0	0	0	3	0
00111	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0
01000	0	0	0	0	0	0	0	0	0	0	45	0	53	0	5	0
01001	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
01010	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0
01100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0
10000	271	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
10001	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10100	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0
10110	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0
11000	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0

Out/In	10000	10001	10010	10011	10100	10101	10110	10111	11000	11001	11010	11011	11100	11101	11110	1111
00000	1,942	0	7	0	21	0	3	0	233	0	0	0	16	2	1	0
00001	24	56	0	0	0	0	0	0	0	15	0	0	0	1	0	0
00010	0	0	30	0	0	0	2	0	0	0	37	0	0	0	5	0
00100	0	0	0	0	195	0	1	0	0	0	0	0	277	12	0	0
00101	0	0	0	0	2	0	0	0	0	0	0	0	0	32	0	0
00110	0	0	0	0	0	0	81	0	0	0	0	0	0	0	136	0
00111	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
01000	0	0	0	0	0	0	0	0	55	0	0	0	5	0	0	0
01010	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
01100	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0
01110	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0
10000	0	0	19	0	291	0	14	0	172	0	0	0	4	0	0	0
10001	55	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

Table 3. (Continued)

Out/In	10000	10001	10010	10011	10100	10101	10110	10111	11000	11001	11010	11011	11100	11101	11110	1111
10010	0	0	0	0	0	0	22	0	0	0	4	0	0	0	2	0
10100	0	0	0	0	0	0	14	0	0	0	0	0	53	0	0	0
10101	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
10110	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0
11000	0	0	0	0	0	0	0	0	0	31	17	0	100	5	2	0
11001	0	0	0	0	0	0	0	0	2	0	0	0	0	38	0	0
11010	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0
11100	0	0	0	0	0	0	0	0	0	0	0	0	0	95	27	0
11101	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0
11110	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4

Notes: Columns in the 5-digit codes of the heading line correspond to the following PEIA modes: GTA, STA, DTT, BIT, CUA; GTAs are Goods Trade Agreements; STAs are Services Trade Agreements; DTTs are Double Taxation Treaties; BITs are Bilateral Investment Treaties; and CUAs are Currency Union Agreements which include unions with one currency as well as currency pegs.

The table includes only rows with at least one positive entry.

BITs obsolete (e.g. the European Union). Some countries adopt unilateral tax exemption for foreign-earned incomes so that concluding DTTs for reasons of double taxation relief becomes an idle strategy. Among the considered PEIAs, DTTs, CUAs, GTAs, and GTAs in combination with STAs are frequently used. Moreover, DTTs are often combined with BITs. Also, GTAs and STAs are frequently combined with DTTs. One of the five PEIAs is chosen in about 6% of the cases, two types of PEIAs are selected in about 1% of the cases, and three or four modes are combined in less than 1% of the cases.

The table also provides information about switching into and out of such integration strategies in the two columns on the outer right. For instance, the table suggests that in 2,405 cases a country-pair gave up on some form of facilitated market access vis-à-vis each other and did not have any PEIA afterwards. Most frequently pairs adopted DTTs – either DTTs only or some combination with other PEIAs (in particular, with BITs). In 3,165 cases country-pairs switched into a combination with just a DTT; 739 of the included pairs adopted BITs without any other PEIA. Over the observation period, 2,021 pairs adopted GTAs alone, 1,379 pairs STAs alone, and 1,055 units adopted some form of CUA without any other PEIA. Moreover, country-pairs most frequently gave up on CUAs alone and switched into some other form of PEIAs. Countries also abandoned a situation with only a DTT or only a GTA quite frequently relative to other forms of PEIAs. Hence, while country-pairs often concluded CUAs or DTTs alone, they also frequently switched out of such PEIAs.

While Table 2 sheds light on the frequency of switching into and out of PEIAs, it provides not much insight into transitions between states of preferential integration. For the five modes of PEIAs (as in Table 2), typical transitions between 1960 and 2005 are identified in Table 3. The table includes the treatments country-pairs may switch out of in rows and treatment states country-pairs may adopt in the columns. Notice that we suppress the information about stayers in this transition matrix and focus on changes of treatment status (due to space constraints, we also suppress all combinations which have not been left once, i.e. lines including zeros only). The top row of Table 3 reveals that – consistent with the top row of Table 2 – many country-pairs leave the state of having no PEIA whatsoever. From that state, country-pairs switch particularly often into one with DTTs only, but they also frequently switch into CUAs only, STAs only, or BITs only. Notice that the sum of all entries from the first rows of the upper and lower panels of Table 3 corresponds to the number of switchers out of the no-PEIA treatment in the first row and last column of Table 2. Moreover, the sum of all cells in the first column of the upper panel in Table 3 corresponds to the number of switchers into the no-PEIA treatment in the first row and penultimate column of Table 2, etc. Table 3 allows us to identify typical integration paths. For example, when country-pairs have a DTT (BIT), the next typical step is to conclude a BIT (DTT). In total, 291 country-pairs with a GTA only concluded a DTT in a next step. Out of a situation with a DTT only, 277

Table 4. Dynamic behaviour (2–4 switches, most frequent cases)

2 switches					
Cases	Initial	Switch 1	Switch 2		
338	00000	00001	00000		
287	00000	00010	00110		
260	00000	00100	00110		
221	00000	10000	00000		
215	00000	10000	10100		
162	00000	00100	11100		
142	00001	00000	00100		
133	00000	00100	10100		
124	00000	10000	11000		
3 switches					
Cases	Initial	Switch 1	Switch 2	Switch 3	
84	00000	00100	00110	11110	
75	00001	00000	00001	00000	
43	00000	00100	11100	11101	
43	00000	00010	00110	11110	
36	00000	00100	00110	10110	
27	00000	00001	00000	01000	
25	00000	10000	10100	11100	
25	00000	00010	00110	00111	
24	00000	10000	11000	11100	
4 switches					
Cases	Initial	Switch 1	Switch 2	Switch 3	Switch 4
20	00000	00001	00000	00001	00000
12	00001	00000	00001	10001	10000
8	00000	10000	10100	11100	11110
8	00000	10000	11000	11010	11110

Notes: 5-digit codes correspond to the following PEIA modes: GTA, STA, DTT, BIT, CUA; GTAs are Goods Trade Agreements; STAs are Services Trade Agreements; DTTs are Double Taxation Treaties; BITs are Bilateral Investment Treaties; and CUAs are Currency Union Agreements which include unions with one currency as well as currency pegs. Cases denotes the number of country-pairs exhibiting a particular integration pattern.

country-pairs concluded a GTA in conjunction with an STA in a next integration step.

Although Table 3 already reveals interesting integration patterns, one may provide even deeper insights into typical paths of integration since 1960. Table 4 summarizes PEIA data for country-pairs which have switched more than once. The top panel of Table 4 shows that if country-pairs switched twice over the whole time-span 1960–2005, they typically did so from a situation without any PEIA into one with a BIT or a DTT to one with both a DTT and a BIT. Of course, the latter is consistent with and could be read off Table 3 with some effort. From 1960 to 2005, this happened in 481 (260 + 221) cases. The other panels of Table 4 reveal that in many cases a DTT or a BIT was just the first step towards deeper integration, where, starting from no agreement, up to four integration agreements have been concluded (e.g. in 84 cases with 3 switches).

Table 3 and Table 4 do not only reveal typical patterns of integration paths since 1960, they also indicate that causal inference of treatment effects in a dynamic analysis is difficult due to small numbers of cases for most transitions.[10] Conversely, we should be aware that estimated effects of individual PEIAs estimated in the past quite likely attributed effects to a single mode where in fact they accrued to a mix (or an integration path) of policies.

4. MULTIPLE TREATMENT EFFECTS OF PEIAS

This paper utilizes propensity score matching to reduce (if not avoid) the self-selection bias of country-pairs into multiple treatments associated with combinations of PEIA modes. We outline technical details and the assumptions in more formal terms in Appendix A. However, it is useful to briefly justify the choice of the approach here. In general terms, propensity score matching aims at eliminating the treatment effect bias from self-selection of cross-sectional units (such as individuals, firms, regions, countries) into some treatments. The approach is most commonly applied in labour economics and development economics to study the effects of specific programs (as binary treatments) on individuals or firms at the micro level. Matching invokes the so-called conditional independence assumption where unconditional mean independence does not hold. Whenever units self-select into a treatment, the unconditional comparison of average outcome for the treated and the untreated should not be interpreted as or is a biased estimate of the true treatment effect. Invoking conditional mean independence through matching implies that we can find a set of observable factors on which we can condition on non-parametrically, so that after conditioning, the self-selection bias is drastically reduced (and ideally avoided). Technically, this is done by estimating the propensity (probability) of receiving the treatment and to select only those treated and untreated units for which comparable propensities (or propensity scores) are found. Then, consistent treatment effects can be estimated from differences between the outcomes of, in this sense, comparable treated and untreated observations. In non-technical terms, two key assumptions for this to work are the following. First, there should be a sufficient overlap in the estimated propensities of the treated and control units so that enough comparable units are available. Second, for similar propensity scores to reflect true similarity between treated and untreated control units, it should be the case that the determinants (observables) underlying the propensity score are similar, too. Otherwise, similarity in propensity scores would be an artifact and the propensity score metric should not be used as a compact metric to establish similarity.[11]

With numerous individuals or firms, these assumptions are often met. However, propensity score matching is also used with aggregate data. Examples in international

[10] If anything, a finer distinction of alternative forms of agreements would only exacerbate that problem.

[11] With all treatment effects estimated in this study, we do not only select comparable units in terms of the propensity score into the control group but also condition on observables as suggested by Blundell and Costa Dias (2009). This eliminates all differences in the first moment of any one of the observables between the control group and the treatment group.

economics are Egger *et al.* (2008) or Baier and Bergstrand (2009), who first used propensity score matching to establish conditional mean independence in trade outcomes (volume or intra-industry trade shares) at the country-pair level. It turns out that even at this level, there is much randomness. For instance, numerous agreements are planned but are never signed for random political, economic, or other reasons. In the time dimension, even more agreements are planned but are not ultimately signed at planned dates. Some agreements are high up at the agenda of one government, but the incumbents lose in subsequent elections, and signing the agreement has low priority with the successor government, etc. All such issues surface in the fact that propensity scores of individual PEIA modes as well as the multiple PEIA modes are determined well but not perfectly by economic, geographical and political fundamentals. In particular, the degree of randomness is non-trivial with as many as 32 possible (and 26 actually used) treatments as underlying this study on the effects of self-selected PEIA modes.[12]

5. DATA

In the context of our analysis, we may distinguish three sets of data: ones underlying the binary indicator variables for PEIA status, observable variables determining multinomial PEIA status, and ones for outcomes that we expect PEIAs to affect.

5.1. Data on PEIA status

We collect data on PEIA status from three sources. First, information about GTAs and STAs stems mainly from the WTO's online database. We augment this information by one about GTAs that have not been notified to the WTO but are included in the datasets of Bergstrand *et al.* (2010). Second, we use information about DTTs and BITs as made available by the UNCTAD online database. Finally, we use the data provided by Reinhart and Rogoff (2004) to classify (*de facto*) currency unions and currency pegs. Since their dataset ends in 2001, we extend it to subsequent years until 2005, using information from the International Monetary Fund (IMF).

5.2. Data on determinants of PEIAs

Our empirical specification of the determinants of PEIAs is similar to the one in Baier and Bergstrand (2004). As mentioned in Section 2, theoretical models of trade and investment liberalization suggest that the same determinants affecting trade also affect foreign direct investment in general equilibrium, and so the same determinants driving

[12] Alternatives to the propensity score matching approach would be difference-in-difference analysis, switching regression, or instrumental variable regression. However, the results in Baier and Bergstrand (2007) and Egger *et al.* (2008) suggest that these alternative methods – though being based on different sets of assumptions (see Wooldridge, 2002) – typically lead to similar results, at least when applied with aggregate country-pair data in international economics.

preferential trade liberalization also drive investment liberalization (see Egger *et al.*, 2006, 2007a, 2007b; or Bergstrand and Egger, 2011).[13] Of course, this does not mean that different types of liberalization are adopted at identical configurations of fundamental observables, but that they depend on the same determinants. The observable variables determining selection into different modes of PEIAs in our model are the following.

5.2.1. Size and factor endowments. We first define $GDP_{ijt} \equiv (GDP_{it} + GDP_{jt})$, where GDP_{it} is country *i*'s real GDP in year *t* and GDP_{jt} country *j*'s real GDP in *t* (in US dollars of the year 2000). Then, the variable $SumGDP_{ijt}$ is defined as $\ln GDP_{ijt}$ for the country-pair *ij*. According to new trade theory, both trade and foreign direct investment are *ceteris paribus* larger among larger countries and so should be the welfare effects of preferential trade and/or investment liberalization. Real GDP figures are taken from the World Bank's World Development Indicators. $SimGDP_{ijt}$ is defined as $\ln[1-(GDP_{it}/GDP_{ijt})^2-(GDP_{jt}/GDP_{ijt})^2]$ and measures two countries' similarity in GDP. *Ceteris paribus*, this variable is supposed to influence a country-pair's propensity to preferentially integrate positively. $SumPOP_{ijt}$ is defined analogous to $SumGDP_{ijt}$, except that we use population numbers POP_{it} and POP_{jt} instead of GDP_{it} and GDP_{jt}, respectively. This variable is not included in previous specifications (see Baier and Bergstrand, 2004), but it seems advisable to use it along with $SumGDP_{ijt}$ if countries differ starkly in terms of their productivity. Population figures are also taken from the World Bank's World Development Indicators. $SimPOP_{ijt}$ is constructed akin to $SimGDP_{ijt}$ and included for the same reason as $SumPOP_{ijt}$. DKL_{ijt} captures the difference in two countries' relative factor endowments, and is measured as the absolute difference in the logarithm of two countries' real GDP per capita (see Egger and Larch, 2008), $|\ln(GDP_{it}/POP_{it})-\ln(GDP_{jt}/POP_{jt})|$. While differences in real per-capita income are an imperfect measure of differences in capital–labour ratios, it is a stylized fact that these measures are highly correlated. Since per-capita income is available for many more countries and much longer time spans than (investment and) capital stock data are, we base DKL_{ijt} on per-capita income rather than capital–labour ratios which would have to be constructed by the perpetual inventory method under numerous assumptions. In models in the vein of Helpman and Krugman (1985) with trade but an absence of multinational firms, trade unambiguously increases with DKL_{ijt} as long as factor price equalization prevails. However, in the presence of multinational firms and if factor price equalization does not apply, DKL_{ijt} still matters for outcome (and the welfare effects of preferential trade and investment liberalization), but its impact is ambiguous (see Bergstrand and Egger, 2007). DKL^2_{ijt} is the squared value of DKL_{ijt} which is included since the impact of DKL_{ijt} on trade, investment, and the welfare effects of PEIAs is inherently non-linear.

[13] In principle, one could include factors such as volatility of real GDP or inflation as determinants of CUAs. We have done so in an extension but the results about estimated propensities are relatively insensitive to that modification.

5.2.2. Geography. *Distance*$_{ij}$ is the natural logarithm of the geographical (great circle) distance between two countries' economic centres. We use the distance variable provided by the Centre d'Études Prospectives et d'Informations Internationales (CEPII). *Common Border*$_{ij}$ is an indicator variable which is unity whenever two countries share a common land border. The variable is taken from CEPII's geographical database too. *Common Language*$_{ij}$ is an indicator variable which is unity whenever two countries share a common (official and other) language. *Same Continent*$_{ij}$ is an indicator variable which is unity whenever two countries are located (at least partly) on the same continent. The latter two variables are constructed on the basis of publicly available data from CEPII. *WTO Member*$_{ijt}$ *(one)* is a variable which is unity whenever only one country in a pair is a member of the GATT or the WTO in year t. *WTO Member*$_{ijt}$ *(both)* is a variable which is unity whenever both countries in a pair are members of the GATT or the WTO in year t (see Egger and Nelson, 2010). *Landlocked*$_{ij}$ *(one)* is an indicator variable which is unity whenever one of two countries exhibits maritime access. *Landlocked*$_{ij}$ *(both)* is an indicator variable which is unity whenever neither one of two countries exhibits maritime access. The latter two variables are taken from CEPII's geographical database.

5.2.3. Politics. *CDW*$_{ijt}$ is a variable which counts the number of days two countries exhibited armed conflict with each other since after World War II. The data are taken from the International Institute for Strategic Studies' Armed Conflict Database. In general, we expect long-lasting armed conflicts to destroy trust and contract viability among business partners from two countries. In this sense, war destroys the basis for trade and, hence, the basis for preferential economic integration. *DYW*$_{ijt}$ counts the number of years since when two countries had the last armed conflict classified as a war with each other. If two countries never had a war, the variable is set to a maximum value of 2005. The data are also taken from the International Institute for Strategic Studies' Armed Conflict Database. We expect the destructive effects of war in trade and foreign direct investment as well as the propensity of preferential integration to be *ceteris paribus* larger for countries which had a recent conflict with each other. *DRD*$_{ijt}$ counts the absolute difference in the number of years two countries' political regimes are in office. The data are taken from Marshall and Jaggers' Polity IV database. Political scientists have found that longer regime durability and, thus, a more stable environment, is prolific to trade so that we expect it to have a positive impact on preferential economic integration. *DPF*$_{ijt}$ measures the absolute difference in the *Polity2* index, which is larger if a country's political freedom is greater. *Ceteris paribus*, we observe that economic activity is larger in countries which display a greater political freedom.

5.2.4. Third-country variables. *Remote*$_{ijt}$ measures the average distance of two countries i and j from all other countries in the sample in a given year t. Defining the great circle distance between countries i and j by D_{ij}, *Remote*$_{ijt}$ is defined as 0,5 $\left\{\ln\left[\sum_{i\neq j} D_{ij}/(N_t-1)\right]+\ln\left[\sum_{j\neq i} D_{ij}/(N_t-1)\right]\right\}$ (see Baier and Bergstrand, 2004;

Egger and Larch, 2008), where N_t is the number of countries in the sample as of year t. $Remote_{ijt}$ varies over time since the sample of countries which are politically independent varies over time. *Ceteris paribus*, a greater remoteness of a pair of countries increases the importance of the two for each other so that we would expect it to display a positive impact on the propensity to grant each other preferential economic market access. $DRKL_{ijt}$ measures the average difference in relative factor endowments of pair ij together with all other countries in the sample in a given year t. Akin to DKL_{ijt}, it is based on the absolute log difference in real per-capita incomes. Finally, to capture general equilibrium effects, we include the share of each type of PEIA in $t-1$ relative to the total number of observations. To be precise, we define $WGTA_{ijt-1} = \sum_{ijt-1}^{N} GTA_{ijt-1} - GTA_{ijt-1}/(N-1)$ (accordingly, $WSTA_{ijt-1}$, $WDTT_{ijt-1}$, $WBIT_{ijt-1}$, and $WCUA_{ijt-1}$).

Standard descriptive statistics of the mentioned observables are provided in Table 5. While we will not engage in calculating and interpreting marginal effects of the observables in the multivariate probit models for the sake of brevity, information about the means and standard deviation of the included fundamentals of PEIA membership will help the interested reader to compare the magnitudes of standardized coefficients relative to each other.

5.3. Data on outcomes

We consider effects of PEIA-related treatments as summarized in Table 2 on two margins each of three outcome variables: nominal bilateral exports of goods in US dollars, nominal bilateral exports of services in US dollars, and nominal stocks of bilateral foreign direct investment. The two margins considered are the *intensive* bilateral (country) margin of activity, where we transform each of the three variables logarithmically and only focus on log (i.e. approximately percentage) changes in outcome, and the *extensive* bilateral (country) margin of activity, where we focus on changes in the propensity to trade or invest directly at the country-pair level.[14] Bilateral goods exports panel data for 137 countries and 46 years are from the United Nations' Comtrade Database, bilateral services export panel data for 129 countries and 14 years are from Francois *et al.* (2009), and panel data on bilateral stocks of outward FDI among 137 countries in 36 years are taken from UNCTAD's Foreign Direct Investment Database.

Table 6 provides some basic descriptive statistics for all considered outcomes and margins of activity. Since services trade and FDI data are only available for fewer countries (country-pairs) and years than goods trade is, the number of observations covered for those two outcomes is smaller than for goods trade.[15] While almost 54%

[14] The results in Egger *et al.* (2011) suggest that these two margins (at least for goods trade) can be analysed as two separate parts of an integrated model.

[15] Notice that we eliminate all countries and years from the data which do not report any outcome to or from any country in a given year. Hence, only countries that are potential sources or recipients of outcome are included.

Table 5. Descriptive statistics for determinants of PEIA membership

	Mean	Std. Dev.	Min.	Max.
$SumGDP_{ijt}$	11.322	1.562	5.805	16.563
$SimGDP_{ijt}$	−1.178	1.477	−9.385	0.693
$SumPOP_{ijt}$	10.731	1.54	5.063	16.563
$SimPOP_{ijt}$	−0.938	1.238	−8.819	0.693
DKL_{ijt}	1.236	0.886	7.37E-06	4.619
DKL^2_{ijt}	2.313	2.846	5.43E-11	21.333
$Remote_{ij}$	8.707	0.329	6.37	9.704
$DRKL_{ijt}$	1.031	0.484	0.002	3.006
$log\ Distance_{ij}$	8.695	0.758	4.088	9.894
$Common\ Border_{ij}$	0.025	0.155	0	1
$Common\ Language_{ij}$	0.152	0.359	0	1
$Same\ Continent_{ij}$	0.256	0.437	0	1
$WTO\ Member_{ijt}$ (*one*)	0.455	0.498	0	1
$WTO\ Member_{ijt}$ (*both*)	0.382	0.486	0	1
$Landlocked_{ij}$ (*one*)	0.301	0.459	0	1
$Landlocked_{ij}$ (*both*)	0.034	0.18	0	1
CDW_{ijt}	13.964	304.672	0	15,389
DYW_{ijt}	1996.139	131.258	4	2,005
DRD_{ijt}	23.441	24.129	0	101
DPF_{ijt}	7.798	6.536	0	76
$WGTA_{ijt_{-1}}$	0.042	0.027	0.004	0.107
$WSTA_{ijt_{-1}}$	0.03	0.033	0.001	0.095
$WDTT_{ijt_{-1}}$	0.064	0.048	0.005	0.171
$WBIT_{ijt_{-1}}$	0.02	0.023	1.00E-04	0.067
$WCUA_{ijt_{-1}}$	0.033	0.022	0.012	0.075

Notes: 641,620 observations; 137 countries$_i$; 137 countries$_j$.

Table 6. Descriptive statistics for outcome variables

	Mean	Std. Dev.	Observations
Goods trade	0.68	3.343	548,713
P(goods trade>0)	0.54	0.498	1,015,498
Services trade	3.507	2.487	23,568
P(services trade>0)	0.728	0.444	32,342
FDI stocks	3.272	3.431	49,590
P(FDI stocks>0)	0.234	0.423	211,680

Notes: All years and available country-pairs. Observations differ across outcomes due to the availability of goods trade (from United Nations' World Trade Database), services trade (from OECD's Online Services Trade Database), and stocks of foreign direct investment (from UNCTAD's Foreign Direct Investment Statistics Online Database). Goods trade, services trade, and FDI stocks are defined in logs.

of the country-pairs display positive goods exports in the suitable subset of observations, FDI stocks are positive in only slightly more than 23%. Services trade is positive in about 73% of the suitable subsample of observations.

Figure 2 summarizes the average probability of positive goods exports (upper left panel), services exports (central left panel), and stocks of outward FDI (lower left panel), as well as the average level (in logs) of positive outcomes (the respective panels on the right-hand side). Using the same logic as in Table 1, the order of the bars in the panels corresponds to: *None*, *GTA*, *STA*, *DTT*, *BIT*, *CUA*; and combinations thereof: *GTA-STA*, *GTA-DTT*, *GTA-BIT*, *GTA-CUA*, *STA-DTT*, *STA-BIT*, *STA-CUA*, *DTT-BIT*, *DTT-CUA*, and *BIT-CUA*.

The figure suggests, for example, that positive goods trade flows (in logs) with a GTA (and no other PEIA) exceed goods trade flows (in logs) among countries without any PEIA (the outcome taking value 1.329 for *GTA* compared with 0.090

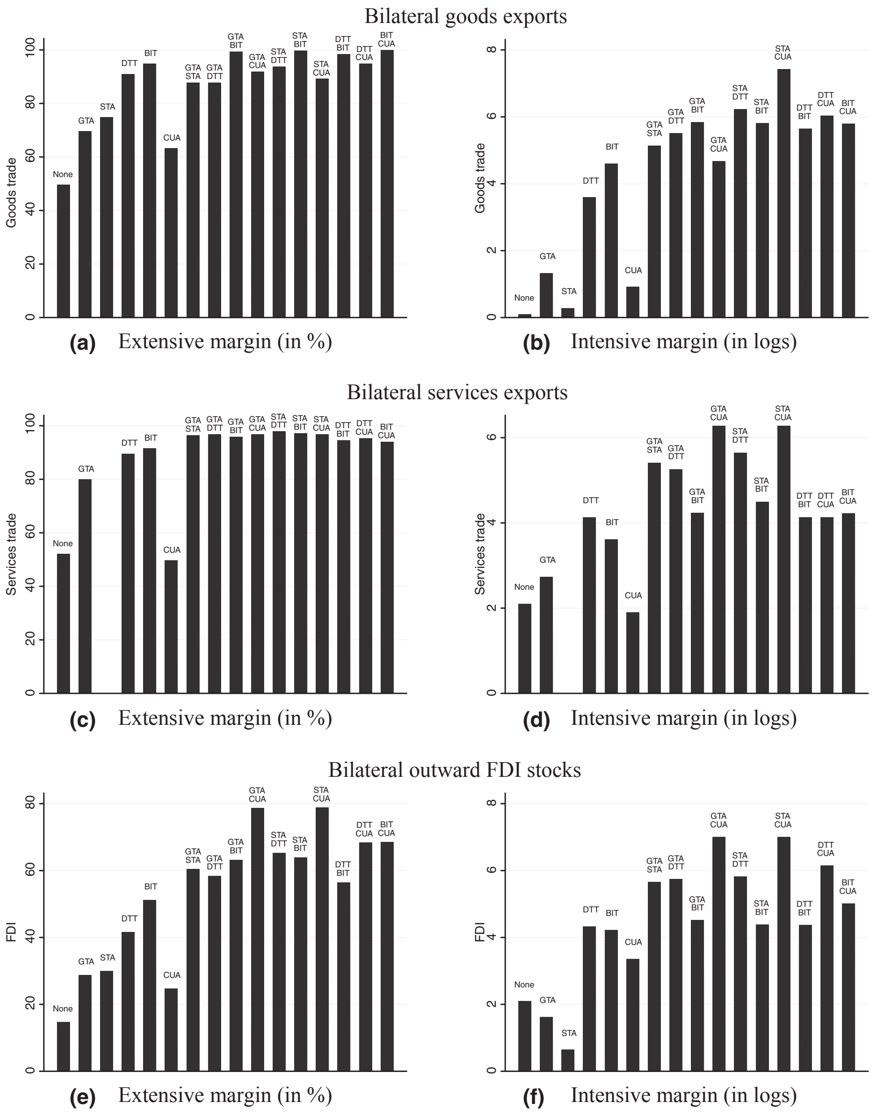

Figure 2. Outcomes inside and outside of PEIAs

for *None*). In fact, bilateral goods exports are lowest when no PEIA was concluded at all. However, positive goods exports tend to be greater with DTTs or BITs alone than with GTAs alone. Similarly, the probability of positive exports is lowest without any PEIA as compared to any alternative with one or more PEIAs. Moreover, multimodal PEIA policies are associated with higher goods exports levels and higher probabilities of positive goods exports than unimodal ones (to see this, compare the last ten bars of the panels which correspond to combinations of PEIAs).

Similar patterns arise for services trade and FDI. For instance, both the level of positive outcome as well as the probability of a positive outcome in the absence of any PEIA (the first bar in the respective graph) tend to be dominated by unimodal or multimodal preferential policies (the other bars to the right of the *None* bar). Moreover, multimodel PEIA strategies (the last ten bars in each panel) tend to be associated with higher levels and a higher probability of positive outcome than unimodal strategies. Also services trade and FDI tend to display higher levels and probabilities of positive outcome with unimodal DTTs or BITs than with unimodal GTAs or STAs. However, for all three outcomes considered, the highest level or probability of positive outcome is associated with some multimodal PEIA that involves GTAs or STAs.

While Figure 2 is suggestive of a positive nexus between different PEIA strategies and goods trade, services trade, and FDI, it does not permit causal statements. The reason for the latter is that countries with particular characteristics – which themselves cause outcome responses – select into specific modes of PEIAs so that the unconditional mean comparisons of outcomes are prone to self-selection bias. The aim of the subsequent section is to avoid that self-selection bias by means of matching methods which estimate unobservable counterfactual outcomes by conditioning on a set of observables so as to improve comparability between country-pairs with alternative PEIA status and reduce if not avoid the self-selection bias.

6. ESTIMATION RESULTS

This section is structured as follows. First, we present the results of (non-linear probability) selection models into multinomial PEIA status. These models provide estimates of the self-selection propensities of country-pairs into all observed states of PEIA treatment in the data. Second, we use those propensities to compare outcomes of country-pairs with different PEIA status to each other and estimate a matrix of treatment effects, comparing different integration options with each other. We will provide a comprehensive set of results for long-run responses to PEIA treatment in 2005 and – for data reasons – summarize short-run responses (within five years after treatment) only for the five most frequent transitions in the data according to Table 3.

6.1. Estimates of multivariate probit models for self-selection into PEIAs

In this subsection we present the results from a multivariate panel probit model based on 5 modes of PEIAs. In contrast to a 5-nomial logit model, this model does neither require that the alternatives are mutually exclusive nor does it impose the assumption of an independence of the choices taken of irrelevant alternatives. In contrast to a 26-nomial logit model, it does not require estimating $26 \times K$ but only $5 \times K$ parameters on covariates. The purpose of this model is to capture the incentives to sign and implement (or net utility associated with) individual PEIAs and combinations thereof. We would expect PEIAs to be signed where the projected gains exceed the corresponding costs.[16]

Table 7 summarizes the coefficient estimates for such a model which is pooled across all years and includes fixed country-pair effects. After deleting observations for which the observable variables are missing, we are left with 641,620 observations. Notice that the multivariate probit model does not ensure that the estimated probabilities sum up to unity over the 26 *taken PEIA modes* but they sum up over the 32 *possible PEIA modes*. Hence, those probabilities have to be adjusted *ex post* so as to sum up to unity for the 26 relevant modes.

The literature proposes a number of ways to assess the goodness of fit of these models. In the sciences, it is common to use Matthews' correlation coefficient (MCC) for doing so. The latter chooses a suitable cut-off value to balance the alpha and beta errors of predicting the binary PEIA modes. Denoting true positive predictions (i.e. a status of one of the respective binary indicator variable) by TP, true negative predictions by TN, and false positive and false negative predictions by FP and FN, respectively, Matthews' correlation coefficient for PEIA mode m is defined as

$$MCC_m = 100\frac{TP_m \cdot TN_m - FP_m \cdot FN_m}{\sqrt{(TP_m + FP_m)\cdot(TP_m + FN_m)\cdot(TN_m + FP_m)\cdot(TN_m + FN_m)}}$$

Notice that $TP_m + TN_m + FP_m + FN_m$ corresponds to the number of observations for any mode m. MCC_m is a compact measure of goodness of fit which is bounded in unitary space. Moreover, it provides guidance for selection of a proper probability

[16] Notice that for many of the included variables in the multivariate probit models we would associate higher values of those variables with bigger net gains from concluding a PEIA (log distance is one exception from this). However, the estimated models are non-linear in the parameters. This entails that marginal effects of individual determinants depend on the values of other variables so that the models implicitly include main effects and also interaction effects. In terms of the specification this means the following with a negative parameter on log distance and a positive one on joint country size. First, a bigger distance or a smaller joint country size between two countries reduces their propensity of signing a specific type of PEIA. However, a bigger distance should *ceteris paribus* be more detrimental for PEIA adoption among larger than among smaller countries. Technically, the reason for the latter is that any type of PEIA happens at relatively low probability, and the marginal effect of any determinant is small, then. However, a larger joint country size for two economies raises that probability *ceteris paribus*. Since the marginal effect of distance rises *ceteris paribus* with a higher probability of any PEIA, larger distances should be more harmful for larger countries.

Table 7. Multivariate treatment model

	GTA	STA	DTT	BIT	CUA
$SumGDP_{ijt}$	0.002	1.042***	0.479***	0.457***	0.113*
	(0.061)	(0.080)	(0.059)	(0.095)	(0.063)
$SimGDP_{ijt}$	0.209***	0.445***	0.302***	0.309***	−0.031
	(0.039)	(0.058)	(0.045)	(0.091)	(0.047)
$SumPOP_{ijt}$	0.462***	0.394***	−0.124*	−0.079	−0.126*
	(0.069)	(0.092)	(0.065)	(0.107)	(0.068)
$SimPOP_{ijt}$	−0.793***	−1.141***	−0.503***	−0.136	−0.155**
	(0.089)	(0.114)	(0.072)	(0.130)	(0.067)
DKL_{ijt}	−0.155***	−0.103	0.112**	0.497***	−0.027
	(0.048)	(0.082)	(0.056)	(0.128)	(0.065)
DKL^2_{ijt}	0.017	−0.037	−0.118***	−0.288***	−0.023
	(0.016)	(0.028)	(0.018)	(0.050)	(0.021)
$Remote_{ij}$	0.126	5.299***	−5.580***	−7.075***	−6.434***
	(0.360)	(0.710)	(0.403)	(0.782)	(0.534)
$DRKL_{ijt}$	−0.037	1.122***	−0.567***	−1.358***	−0.056
	(0.054)	(0.099)	(0.064)	(0.133)	(0.059)
$log\ Distance_{ij}$	−0.353***	0.115***	−0.511***	−0.449***	−0.137***
	(0.026)	(0.028)	(0.025)	(0.035)	(0.026)
$Common\ Border_{ij}$	0.359***	0.286***	−0.268***	−0.241**	0.142
	(0.068)	(0.090)	(0.088)	(0.117)	(0.088)
$Common\ Language_{ij}$	0.273***	−0.084*	0.188***	0.185**	0.831***
	(0.037)	(0.045)	(0.047)	(0.074)	(0.028)
$Same\ Continent_{ij}$	0.239***	0.368***	−0.093**	−0.331***	0.237***
	(0.038)	(0.042)	(0.040)	(0.057)	(0.036)
$WTO\ Member_{ijt}$ (one)	0.446***	0.859***	0.256***	0.363***	0.103***
	(0.047)	(0.079)	(0.052)	(0.102)	(0.036)
$WTO\ Member_{ijt}$ (both)	0.701***	1.219***	0.490***	0.394***	0.108***
	(0.053)	(0.084)	(0.058)	(0.110)	(0.042)
$Landlocked_{ij}$ (one)	−0.002	−0.763***	0.124***	0.254***	−0.056*
	(0.038)	(0.046)	(0.038)	(0.052)	(0.033)
$Landlocked_{ij}$ (both)	−0.130*	−1.427***	0.305***	0.394***	−0.059
	(0.077)	(0.113)	(0.111)	(0.119)	(0.103)
CDW_{ijt}	−3.00E-05	−2.00E-05	−2.00E-04**	−1.00E-04***	2.00E-05
	(5.00E-05)	(4.00E-05)	(6.00E-05)	(4.00E-05)	(4.00E-05)
DYW_{ijt}	1.00E-04	−1.04E-06	−1.00E-04	−2.00E-04**	−7.00E-05
	(8.00E-04)	(1.00E-04)	(9.00E-05)	(9.00E-05)	(1.00E-04)
DRD_{ijt}	−0.006***	−0.008***	0.008***	0.006***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
DPF_{ijt}	−0.009***	−0.005**	−0.011***	−0.017***	−0.007***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
$WGTA_{ijt_{-1}}$	7.493***	−12.486***	5.730***	5.929***	1.700***
	(0.347)	(0.565)	(0.432)	(1.020)	(0.591)
$WSTA_{ijt_{-1}}$	−3.064***	10.935***	−7.672***	−8.977***	−3.708***
	(0.363)	(0.993)	(0.476)	(1.116)	(0.885)
$WDTT_{ijt_{-1}}$	−7.707***	−5.440***	−4.854***	−17.031***	−5.141***
	(0.603)	(1.042)	(0.661)	(1.226)	(0.816)
$WBIT_{ijt_{-1}}$	16.177***	−29.929***	11.432***	26.585***	14.385***
	(0.995)	(2.133)	(1.093)	(1.948)	(2.061)
$WCUA_{ijt_{-1}}$	−14.517***	5.710***	1.171**	4.147***	10.022***
	(0.882)	(1.002)	(0.484)	(1.221)	(0.523)
MCC	69.21	69.51	68.94	70.66	68.63
	(0.68)	(0.73)	(0.70)	(0.58)	(0.69)

Notes: 641,620 observations. Multivariate probit model. Pair-specific means and year trends included. ***, **, and * indicate that coefficients are significantly different from zero at 1, 5, and 10%, respectively. Robust and clustered (by country-pair) standard errors in parentheses. MCC is the maximum attainable Matthews' correlation coefficient of the estimated models. The values reported materialize at the probability cut-off values reported in parentheses below.

cut-off value, at and above which we associate probabilities as to suggest a prediction of a binary indicator value of one (positive) and below which we associate them with a zero value (negative). At the cut-off value which maximizes MCC_m, the level of MCC_m indicates the goodness of fit of the estimated models, akin to an R^2 in linear regressions. For computation of MCC_m, we have to employ the *m*-specific vector of estimated probabilities for selection into the *m*th PEIA mode. Similar to the parameter estimates, those probabilities have to be simulated by Monte Carlo methods. The corresponding MCC coefficient for each mode and the probability cut-off level at which it materializes (the latter in parentheses) are given at the bottom of Table 7.[17]

The findings with regard to self-selection into PEIA modes can be summarized as follows. First, the values of MCC_m are quite high. In general, the predictive power is lowest for CUAs, and it is highest for BITs. Consistent with the great variability of the frequency of events across PEIA modes, the highest attainable predictive power as captured by MCC_m is reached at rather different mode-specific probability cut-off values. However, the measure is rather stable in a fairly large neighbourhood around the optimal cut-off level. For instance, MCC_m does not change by more than one percentage point within a symmetric interval of about 30 percentage points in probability space around the optimum cut-off probabilities. Second, many of the coefficients on observables are significantly different from zero at conventional levels. This suggests that the included observables are relevant determinants of self-selection into PEIA modes. When considering parameter point estimates, many of them are in line with earlier results from the literature considering univariate selection into PEIA modes: most of the observables included in the PEIA selection models would take on the same parameter signs in (reduced-form) gravity models of bilateral goods exports, services exports, or stocks of outward FDI.[18] The latter suggests that those countries which have stronger trade or foreign investment relationships also tend to integrate more likely in PEIAs of one or the other form. This is in line with the argument that the

[17] We provide a table including correlation coefficients for the disturbances across PEIA modes in a separate Web Appendix. We ran various alternative multivariate selection models to the one in Table 7. For instance, we ran models that distinguished between deep and shallow GTAs, using the classification introduced by Horn *et al.* (2010). Following their work, the WTO classified numerous GTAs accordingly, which would allow for many more types. However, this leads to more and to smaller treatment and control groups so that the treatment effect estimation becomes the more imprecise, the more such types one distinguishes. Similarly, one could distinguish between PEIAs prior to the mid-1990s and after that since many GTAs concluded in the latter period contain non-goods-trade-related provisions (see also Baldwin, 2011). Moreover, in an alternative set of models we classified BIT as unity whenever a GTA included investment provisions. Finally, we estimated dynamic multivariate probit models, following Wooldridge's (2005) approach towards the modelling of initial conditions. The associated results cannot be presented here for reasons of space constraints, but they are included in a Web Appendix.

[18] Notice that the results for the time-variant variables are difficult to compare with earlier work. For instance, we include measures of country size in terms of GDP as well as population. Hence, the measures relating to GDP, conditional on population size, reflect an impact of total (average) and relative (similarity of) productivity, while population reflects country size itself. Other authors such as Baier and Bergstrand (2004) did not include both measures together, but the parameter estimates suggest that either of them should be included in the empirical model.

benefits from integration are larger among natural integration partners, and it suggests that one should be careful with drawing firm quantitative conclusions from unconditional mean comparisons as from Figure 2, since they are likely masked by a self-selection bias. Third, the means of the time-variant variables (not reported) are jointly significant at 1% in any one of the equations. This indicates that there are time invariant determinants PEIA integration is correlated with, and by pooling the selection models across time without taking care of the time-invariant unobservable variables runs at risk of parameter bias. Fourth, there is a significant correlation between the disturbances across PEIA modes, on average. Thus, by modelling selection into alternative PEIA modes as independent processes, one incurs a risk of parameter bias.

The predicted probabilities of adopting any one out of the five PEIA modes can now be used to predict the response probabilities of all 26 used options of combining those modes, denoted p^{s}_{ijt} for selection of country-pair ij in year t into treatment s (see the Appendix for more details). As said before, to ensure that the estimated probabilities sum up to unity across all actually used options, we rescale them by distributing the estimated ones of unused treatments across all used treatments proportionately to the relative probability mass. Let us denote rescaled probabilities of selecting into treatment s as $\tilde{p}^{s}_{ijt}$. We may use $\tilde{p}^{s}_{ijt}$ to estimate the unobserved outcome for counterfactual treatment, say, $\tilde{s}$, and determine the treatment and control groups based on $\tilde{p}^{s}_{ijt}$ and $\tilde{p}^{\tilde{s}}_{ijt}$ for the subsample of observations which actually received treatment s or $\tilde{s}$ as indicated in the Appendix.

6.2. Estimates of 'long-run' treatment effects of PEIAs in 2005

We enforce three conditions for the quality of matching – that is, the construction of the control group: first, there is common support in the domain of $\mathbf{p}^{s} = (p^{s}_{ijt})$ for units that received treatment s versus ones that received, say, $\tilde{s}$; second, we only use treated and control observations within a radius of 1% in $\mathbf{p}^{s}$-space for matching; third, after matching based on $\mathbf{p}^{s}$, the comparison between units with treatment s versus $\tilde{s}$ is not driven by first-moment differences in any observables determining treatment status; finally, we focus on treatment effects where at least 50 observations are available. Hence, we will consider fewer treatment-control comparisons than is possible from Table 2.

To assess long-run effects of PEIAs on outcome, we focus on data of 2005. In this sense, matching ensures that treated observations of 2005 are matched onto ones of the same year, but we disregard any time structure in the acquisition of treatment status. Since the high number of treatment-control-group combinations implies that hundreds of possible treatment effects could be estimated, let us primarily focus on treatment effects using as a control group those units with a no-PEIA treatment. The matching results for average treatment effects of the treated (ATTs) are presented in Table 8. Each column refers to one of the six outcomes. The results show a very clear pattern: more integration (compared with no integration at all) of any kind is associated with a positive treatment effect.[19]

One interesting finding in Table 8 is that the extensive margin of bilateral activity is often not affected by any kind of integration. However, we find effects on the extensive margin of goods exports when country-pairs enter a GTA- or a GTA cum DTT-type of integration (Treatments 10 and 12), which seems plausible. With regard to the intensive margin of goods exports, treatment combinations involving a BIT lead to particularly large treatment effects. The largest ATT, though, is estimated for country-pairs which concluded every type of PEIA but a CUA (last treatment combination in Table 8). The findings for services exports confirm for both margins that BITs as well as DTTs, both aiming at the activities of multinational firms, have effects on bilateral economic activity that go beyond the bilateral activity of multinational firms. At the intensive margin of services exports, we estimate the largest ATT (1.30) for a combination involving three PEIA modes (a GTA, an STA, and a DTT). Finally, we find that the extensive margin of bilateral FDI is negatively affected when a DTT is combined with a BIT or when a DTT is combined with a GTA. While both findings seem counterintuitive, we should note that DTTs not only include double taxation relief for multinational firms, but also facilitate information exchange between countries so that tax avoidance becomes more difficult for multinationals and, thus, FDI less attractive.

As for the intensive goods exports margin, the largest effect is found for the combination GTA, STA, DTT, and BIT. The estimated ATT implies that integration in these four dimensions is associated with about 750% more bilateral FDI relative to the counterfactual situation of no integration at all. This quantitatively significant effect highlights the huge long-run benefits of integration in general and (goods and services) trade and investment liberalization in particular.

Figure 3 summarizes all estimated (and significant) ATTs for all possible treatment-control combinations. For each outcome, we report two standard box plots: the left (right) one always corresponds to effects for which the number of any kind of PEIAs in the treatment (control) group exceeds the number of PEIAs in the control (treatment)

[19] We have replicated Table 8 using a simple OLS framework, unconditionally comparing outcome of the treatment combinations as listed in Table 8 with those of the control group without any PEIA (detailed results are reported in the Web Appendix). A comparison between the unconditional OLS estimates and our basic findings broadly suggests that ATTs are typically overestimated by OLS, while standard errors are typically underestimated. As our matching approach relies on estimated probabilities which may differ according to the specification of the selection model, we also performed rank correlation tests for the models mentioned in n. 17. To do this, we proceeded in three steps. First, we used predicted probabilities for 32 PEIA modes from different model specifications and ranked these probabilities in ascending order. Second, we calculated pair-wise correlation coefficients for the ranks between the different specifications. Third, we determined the (weighted) average of these correlations (as weights, we used the number of matched country-pairs in Table 8). We found a correlation coefficient of 0.93 between the basic specification and the specification which employs an adjusted measure of BITs, setting it to unity for deep trade agreements which include investment provisions. The correlation coefficient between the basic specification and the dynamic multivariate selection specification is 0.54.

Table 8. Long-run effects of PEIAs (control: no PEIA)

	Goods exports		Services exports		FDI	
	Margin					
Treatment:	Extensive	Intensive	Extensive	Intensive	Extensive	Intensive
0 0 0 0 1	0.067	0.379	−0.105	.	0.224	.
	(0.054)	(0.392)	(0.102)	.	(0.150)	.
0 0 0 1 0	0.079**	2.420***	0.035	0.322	0.131*	1.437*
	(0.040)	(0.317)	(0.097)	(0.454)	(0.080)	(0.765)
0 0 1 0 0	0.078***	1.547***	0.175**	0.695***	0.004	1.133***
	(0.009)	(0.096)	(0.024)	(0.107)	(0.025)	(0.253)
0 0 1 1 0	0.040***	2.228***	0.132**	0.881***	−0.098***	1.548***
	(0.011)	(0.150)	(0.026)	(0.148)	(0.036)	(0.385)
0 0 1 1 1	0.034	2.671***	.	.	0.019	.
	(0.143)	(0.930)	.	.	(0.190)	.
0 1 0 0 0	0.023	0.383***	.	.	−0.025	0.002
	(0.019)	(0.144)	.	.	(0.065)	(0.658)
0 1 0 1 0	0.051	2.559***	.	.	.	.
	(0.090)	(0.686)	.	.	.	.
0 1 1 0 0	0.062	2.009***	.	.	0.275*	.
	(0.056)	(0.590)	.	.	(0.168)	.
0 1 1 1 0	0.048	2.312***	.	.	.	.
	(0.059)	(0.917)	.	.	.	
1 0 0 0 0	0.132***	0.968***	0.059	0.094	0.122**	0.280
	(0.022)	(0.149)	(0.081)	(0.322)	(0.060)	(0.596)
1 0 0 1 0	0.036	1.944*	0.133**	1.003***	.	.
	(0.179)	(1.112)	(0.069)	(0.337)	.	.
1 0 1 0 0	0.109***	1.544***	.	.	−0.116**	.
	(0.024)	(0.221)	.	.	(0.054)	.
1 0 1 1 0	0.020	1.602***	0.050	0.802***	−0.004	.
	(0.042)	(0.414)	(0.071)	(0.394)	(0.114)	.
1 1 0 0 0	0.057	0.193	.	.	0.045	1.525
	(0.041)	(0.286)	.	.	(0.107)	(1.261)
1 1 0 1 0	−0.103	2.883***	.	.	0.281	.
	(0.183)	(1.165)	.	.	(0.229)	.
1 1 1 0 0	0.039	2.764***	0.111	1.303***	0.156**	0.424
	(0.030)	(0.253)	(0.085)	(0.334)	(0.062)	(0.749)
1 1 1 0 1	−0.047	2.792***	0.018	1.24	−0.049	.
	(0.171)	(1.121)	(0.246)	(0.807)	(0.256)	.
1 1 1 1 0	0.054	2.932***	0.052	1.021***	0.187**	2.135***
	(0.037)	(0.357)	(0.093)	(0.411)	(0.089)	(0.740)

Notes: Reported are treatment effects (ATTs) and standard errors (in parentheses) from weighted regressions with at least 50 observations. The control group corresponds to no PEIA at all (0 0 0 0 0). 5-digit codes of the treatments correspond to the following PEIA modes: GTA, STA, DTT, BIT, CUA; GTAs are Goods Trade Agreements; STAs are Services Trade Agreements; DTTs are Double Taxation Treaties; BITs are Bilateral Investment Treaties; and CUAs are Currency Union Agreements which include unions with one currency as well as currency pegs. Extensive margin is the probability of positive bilateral goods exports, services exports, and FDI, respectively. Intensive margin is the log of bilateral goods exports, services exports, and FDI, respectively. Dots denote that coefficients could not be estimated because of insufficient observations.

group, so that we would expect rather positive (negative) ATTs. Denoting the number of PEIAs of the treatment (control) group by TG (CG), we provide information on whether the number of PEIAs is lower or higher in the treatment group relative to the control group ($TG \lesseqgtr CG$) below each figure. Note that Figure 3 includes all

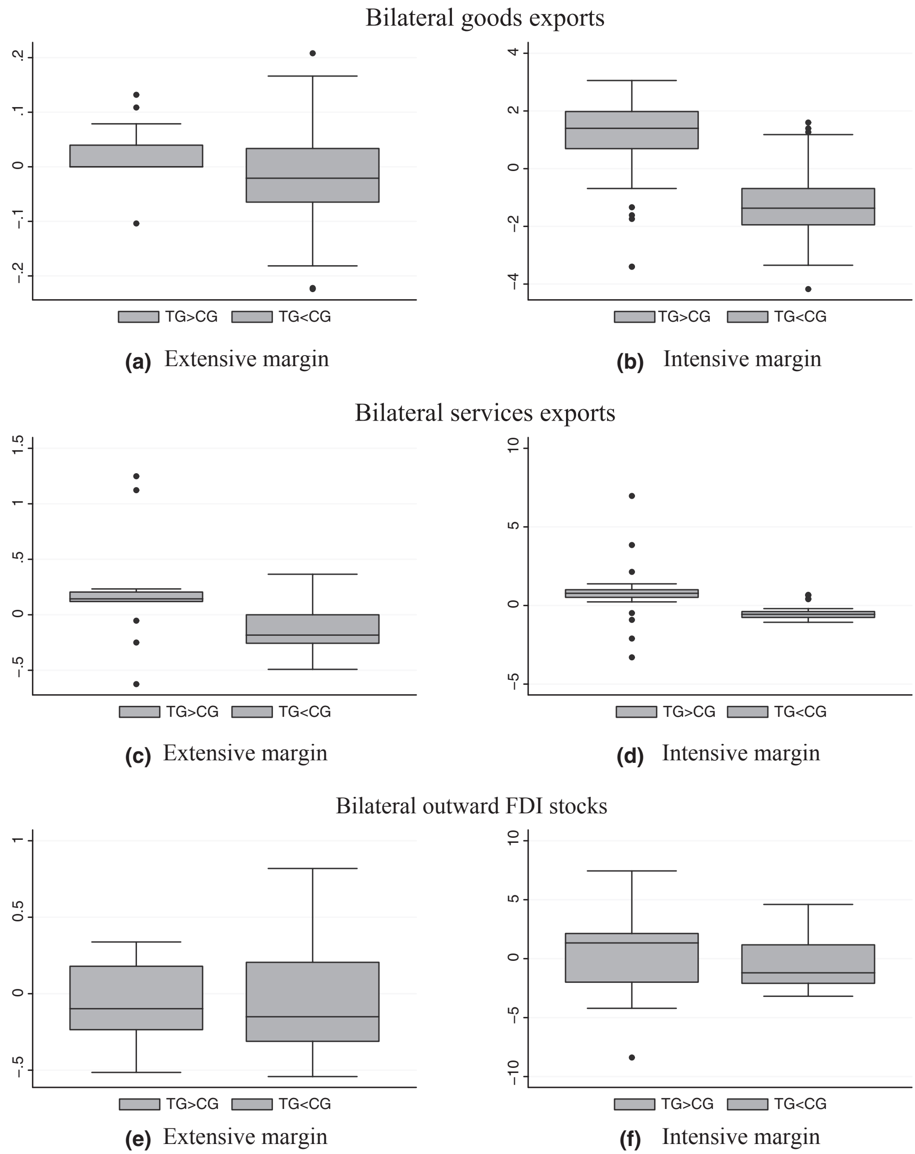

Figure 3. Distribution of (significant) long-run treatment effects

coefficients for which at least 50 observations were available, and which proved significant at the 10% level.[20] The figure suggests that trade (of goods or services) will likely rise if PEIAs are implemented in or added to existing ones. The results for the extensive margin are much less clear-cut. The latter is not surprising since PEIAs are rarely

[20] We provide the same graphs based on all estimated coefficients, including the insignificant ones, in the Web Appendix.

implemented among countries where no cross-border economic activity is going on at all before.

6.3. Estimates of 'short-run' treatment effects of PEIAs

While we did not condition on the treatment state of origin or the timing of treatment adoption in the analysis of long-run ATTs as defined in the previous subsection, we will scrutinize these issues in this subsection. Here, we estimate ATTs conditional on the state of origin so that treated and control units have the same PEIA treatment in an interval of five years prior to treatment of the treated, $\tau \in [-5, -1]$. While the control units stay in the state of origin after that, the treated switch at time $\tau = 0$ and stay in that new state during $\tau \in [0, +5]$ while the control units are required to stay in the state of origin. Outcomes of the treated and the matched control units are then compared over the time path $\tau \in [0, +5]$ as outlined in the Appendix. Hence, in contrast to the long-run effects in the previous subsection, it is now ensured that (i) all treated units switch into treatment at time $\tau = 0$, (ii) that all treated and untreated units each have an identical treatment in any year considered, and (iii) that the treatment between the treated and controls is identical at time $\tau < 0$. Since we consider effects only up to generic year $\tau = 5$, we dub such ATTs short-run effects. Table 9 summarizes those short-run ATTs for the five most frequent treatment switches according to Table 3.

For brevity, we report ATTs of a treatment change in generic period $\tau = 0$ on the extensive country margin of any outcome in period $\tau = 0$ and on the intensive country margin in periods $\tau = 0, \ldots, 5$. The results can be summarized as follows. First, countries display short-run effects in response to changes in treatment relatively rapidly after treatment at both the extensive and the intensive country margin on outcome. This conclusion can be drawn when considering the estimated evolution of ATTs at the intensive margin after new treatment. Moreover, it can be drawn when comparing the results in Table 9 with the corresponding ones in Table 8. For example, while the average response of a country-pair to a switch out of the situation with no PEIA at all into one with just a DTT amounts to about 1.32 immediately after the switch, the corresponding long-run effect shown in Table 8 is about 1.55. Second, the short-run findings broadly confirm the long-run results: the response to DTT integration is strong not only for FDI but also for goods trade; switching out of a situation with integration into a situation without integration is associated with a negative treatment effect; the effects of CUAs are ambiguous; integration by GTA or STA seems to cause negative effects on FDI. The reason for the latter might be, and this could be particularly the case for STAs affecting services trade within multinational firms, that STAs entail agreements on information exchange between countries which restrict the firms' opportunities for tax planning in general and transfer pricing in particular.

Table 9. Short-run effects of PEIAs

Outcome:	Out-in 00000–00100	Out-in 00001–00000	Out-in 00000–10000	Out-in 00000–01000	Out-in 00000–00001
P(goods trade>0)$_{(\tau=0)}$	0.120	−0.623	0.056	0.037	0.056
Goods trade$_{(\tau=0)}$	1.319	−0.326	0.334	0.054	0.352
Goods trade$_{(\tau=1)}$	1.285	−0.423	0.314	−0.031	−0.037
Goods trade$_{(\tau=2)}$	1.364	−0.175	0.393	−0.126	0.001
Goods trade$_{(\tau=3)}$	1.421	−0.079	0.559	−0.085	−0.017
Goods trade$_{(\tau=4)}$	1.434	−0.190	0.518	−0.061	−0.080
Goods trade$_{(\tau=5)}$	1.547	−0.264	0.629	0.018	0.242
P(services trade>0)$_{(\tau=0)}$	0.165	.	0.250	.	.
Services trade$_{(\tau=0)}$	0.209	.	0.500	.	.
Services trade$_{(\tau=1)}$	0.132	.	−0.089	.	.
Services trade$_{(\tau=2)}$	0.182	.	.	.	.
Services trade$_{(\tau=3)}$	−0.016	.	.	.	.
Services trade$_{(\tau=4)}$	0.017	.	.	.	0.087
Services trade$_{(\tau=5)}$	0.198	.	.	.	0.308
P(FDI stocks>0)$_{(\tau=0)}$	0.049	.	−0.080	−0.021	0.074
FDI stocks$_{(\tau=0)}$	0.775	.	−0.658	−2.313	.
FDI stocks$_{(\tau=1)}$	1.044	.	−0.227	−5.756	.
FDI stocks$_{(\tau=2)}$	0.929	.	.	−4.717	.
FDI stocks$_{(\tau=3)}$	0.938	.	.	−1.989	.
FDI stocks$_{(\tau=4)}$	1.438	.	−2.494	−0.509	0.253
FDI stocks$_{(\tau=5)}$	1.268	.	−0.381	−0.085	.

Notes: Reported are average estimated coefficients. 5-digit codes of the heading line correspond to the following PEIA modes: GTA, STA, DTT, BIT, CUA; GTAs are Goods Trade Agreements; STAs are Services Trade Agreements; DTTs are Double Taxation Treaties; BITs are Bilateral Investment Treaties; and CUAs are Currency Union Agreements which include unions with one currency as well as currency pegs. Dots indicate that coefficients could not be estimated because of insufficient observations.

Unfortunately, missing data on outcomes such as services trade and FDI stocks do not permit giving as complete a picture as we wished to.[21] Hence, an econometric analysis of ATT dynamics which relies on the rich pattern of PEIA adoptions as outlined in Tables 2–4 is unfortunately not feasible for many treatments in general and for just a very few treatments in case of services and goods trade. We have to wait for nature to generate more data on this to be able to shed light on the impact of such changes in subsequent work.

8. CONCLUSIONS

Unlike previous research, this paper considers joint selection into and effects of five modes of preferential economic integration agreements (PEIAs): goods trade agree-

[21] Recall that we require that no treatment change occurred at least five years prior to the change investigated in Table 9. This leads to relatively small windows of data available for, e.g. services trade and FDI stocks.

ments (GTAs), services trade agreements (STAs), double taxation treaties (DTTs), bilateral investment treaties (BITs), and currency unions as well as currency pegs (CUAs). In a broad descriptive analysis we unveil typical integration patterns of economies. Using individual PEIAs in isolation appears to be a rare strategy. The considered PEIAs obviously address not only goods trade but also trade in services as well as foreign direct investment. It is a potential shortcoming that virtually all previous research considered only one of those outcomes at a time. Such a strategy may be misleading to the extent that the three mentioned outcomes go hand in hand not only because of the engagement of multinational firms in goods trade (their engagement in services trade is natural due to the supply of headquarters services within multinational networks), but also due to complementarities of such activities by way of information exchange about markets and contracts across (exporting and foreign investing) firms in the same parent country.

Our investigation shows that single and combined PEIAs tend to trigger positive effects not only on single outcome but typically on all outcomes. While GTAs are often concluded in a first integration step, the effects on goods trade with investment liberalization through BITs or DTTs in a first step are typically much larger than when liberalizing goods trade *per se*. Hence, it appears important to consider the interdependence of goods trade, services trade, and foreign direct investment when adopting measures of preferential liberalization. It is not necessarily the case that the most direct PEIA measures available display the largest effects on a primarily targeted outcome of interest. Moving first in (or combining preferential trade liberalization with) investment liberalization promises bigger effects for goods trade than when focusing on liberalizing tariff and non-tariff barriers to goods trade alone. In general, covering several issues (such as trade, services, and investment) tends to benefit each outcome more likely and to a larger extent than when focusing on single-issue liberalization. In part, this is reflected in the reality of today's treaty drafts that often reach out to issues which lie beyond their intended scope (e.g. GTAs include regulations about investment, etc.). Offering a mix of PEIA modes to international business seems to be a viable alternative to such practice.

Our descriptive analysis shows that many country-pairs start out with signing DTTs as one form of liberalizing investment in a broad sense. When comparing effects on outcomes, DTTs and, similarly, BITs – as a second form of preferential investment liberalization – display large effects on outcomes. Thus, in comparison, we should see more BITs in view of their relatively large effects not only on investment but also on goods trade. In other words, the evidence suggests that BITs are under-represented relative to GTAs or DTTs in international economic policy regarding their relatively sizeable effects on economic activity. With respect to sequencing, we would expect relatively more countries to sign BITs relative to GTAs or STAs as a first preferential integration step and, subsequently, to combine them with trade agreements (GTAs or STAs).

Discussion

Thorsten Beck

University of Tilburg

This is an important and policy-relevant paper that explains the adoption of different combination of agreements between countries on preferential market access, including agreements on goods trade, services trade, foreign direct investment, double taxation and currency unions. More importantly, the authors link the different combinations of such agreements to actual trade and FDI between country pairs over time. Compared to the previous literature, this paper takes a more comprehensive look at different agreements, while previous papers have considered specific forms of agreements, be they on trade in goods, trade in services or FDI.

If one wants to set the paper in the broader context of the literature, it links the literature on the determinants of international trade and the literature on what drives preferential economic integration agreements, including political economy determinants. By using a multinomial selection model, the authors gauge both the long-term effects of specific combinations of these agreements and the short-run responses to transitioning from one combination to another (e.g. by adding an additional agreement), while controlling for selection bias and potential endogeneity.

While the paper provides interesting and important insights, I have some conceptual concerns. The first concern refers to the list of agreements the authors have chosen. While taking a much more comprehensive view on integration agreements than the literature typically does, there are other important elements missing that can explain international trade in goods and services and foreign direct investment, including capital account liberalization, labour market integration, visa-free travel and so on. Where does one draw the line and does the limitation to specific agreements result in an omitted-variable bias? Second, what is the effect of preferential economic integration agreements compared to other policies that might explain variation in trade and foreign direct investment? And, finally, are we asking the right question by gauging the relationship between these agreements on trade rather than looking at the additional effect of the agreements on trade beyond what would be the natural level of trade between two countries. I would expect the effect of such agreements to vary across country-pairs that have different levels of natural trade, related to distance, different history etc. By including these determinants of natural trade intensity as determinants of preferential agreements, the authors might overestimate the effect of agreements on trade.

I would like to add some comments on the estimations. Since World War II, there has been a general trend towards global trade agreements, which might have strengthened the effect of preferential agreements between two or more countries. It is not quite clear whether this is properly controlled for. Further, there might be a critical difference between bilateral and multilateral agreements that is not captured in the current analysis.

In summary, this paper presents a fascinating dataset and sets out a very relevant and timely research agenda. I expect additional analysis in future work to disentangle the effects of different preferential economic integration agreements and their sequencing on international trade and investment, including differential impact across different country groupings.

Paola Conconi

Free University of Brussels

This paper uses a rich dataset, covering 210 economies over the period 1960–2005, to examine both selection into and the impact of five modes of preferential economic integration agreements (PEIAs): goods trade agreements (GTAs), services trade agreements (STAs), double taxation treaties (DTTs), bilateral investment treaties (BITs), and currency unions/ currency pegs (CUAs).

The analysis consists of two main steps. First, the authors estimate a multivariate non-linear probability model to examine countries' choices among 26 PEIA configurations (out of the possible 32). Second, they examine the implications of these agreements choices on the probability and the level of bilateral trade in goods and services, as well as the probability and the level of bilateral foreign direct investment (FDI). To deal with concerns of self-selection of country-pairs into different agreements, the authors use propensity score matching, estimating the probability of each country pair of receiving the treatment (i.e. being in one kind of PEIA arrangement) and selecting only those treated and untreated units with comparable propensities. A key assumption to be able to apply this methodology to international agreements is that, after controlling for economic, geographical and political variables driving selection into PEIAs, there should be enough randomness in whether or not governments actually sign an agreement or in when an agreement is signed.

In general, the paper challenges the existing literature assessing the impact of preferential agreements. Previous studies in this literature focus one policy area (e.g. trade agreements), without considering possible interactions with agreements in other policy areas (e.g. tax treaties). Implicitly, these studies assume that the various types of agreements are concluded independently of each other and that they do not influence each other's impact in determining outcome. Egger and Wamser provide evidence of an important overlap in the conclusion of different types of PEIAs, emphasizing the difficulty of examining their impact in isolation from each other.

An interesting finding of this study is that agreements that cover several issues (e.g. trade and investment) are more likely to benefit trade and FDI (both at the extensive and intensive margin) than agreements focusing on single-issue liberalization. This supports the idea that trade and FDI are complements, for example, due to the global fragmentation of production (Head and Ries, 2004) or to information spillovers (within and between) firms serving foreign markets via exports or FDI (Conconi *et al.*, 2012).

The paper focuses on bilateral agreement choices. In reality, most PEIAs are not limited to two countries and strategic interactions play a key role in governments' decisions on whether to sign preferential agreements. Indeed, the results presented in Table 7 show that systemic variables (the number of countries having already signed different types of PEIAs) are the most important determinants of governments' choices, emphasizing the importance of 'contagion effects' (Baldwin, 1993). An interesting avenue for future research is to further explore strategic effects across preferential agreements.

Panel discussion

Andrea Ichino was slightly sceptical about the use of propensity score matching (PSM) in the empirical analysis. He argued that in order to employ this technique the authors should have a clear story explaining why, conditioning on the explanatory variables, the treatment should be random. Ichino also emphasized that the assumption of independence across observations is required for the application of PSM and wondered whether the countries could be considered as independent observations. Regarding the first-stage of the multivariate probit, Isabel Schnabel asked how the authors deal with the persistence of treatments.

Replying to Paola Conconi, Peter Egger said that they also have results on the deepening of trade agreements. In particular, in one regression they distinguish between deep and shallow goods trade agreements. However, Egger warned the panel that disentangling the two also broadens the number of modes and leads to smaller cells, which in turn makes it more difficult to estimate treatment effects. According to Egger, this approach in the extreme results in just case studies and impedes comparison-making. Addressing Schnabel's question, he pointed out that dynamic selection models were estimated (probabilities compared and ranked). Referring to Thorsten Beck's discussion, Egger indicated that a mixture of PSM and weighted regressions was used in the study and that the relevant variables were in addition controlled for in the second step. He reiterated that they always condition on the same observables that enter the selection equation in the second step when estimating the average treatment effects. On the issues of heterogeneous treatment effects and emphasizing contagion, Egger felt that these were good points and said that he would think about them more. He also agreed with Beck that it would be useful to examine other policy issues (labour market, competition policies etc.) but at the same time raised the argument that this would reduce the cell sizes for each of the considered combinations (hence a trade-off). Egger, like Beck, thought that a short-term approach would have been better but said that it was not very feasible (link to the literature would be lost, numbers become small – data generating process not running long enough). Moving on to Ichino's comments, Egger contended that there is enough randomization present to validate the PSM methodology. Specifically, he noted that although countries may have plans to ratify a treaty, they might not when the time comes. Egger concluded the discussion by saying that general equilibrium effects should not be too large as otherwise they would appear at the level of the outcome equation which would violate the main assumption.

APPENDIX A. CONCEPTS AND ASSUMPTIONS FOR ESTIMATION OF MULTIPLE ENDOGENOUS TREATMENT EFFECTS

A.1. SOME NOTATION AND CONCEPTS

The problem at stake is one of 32 mutually exclusive treatment states a generic country-pair $ij \in \{1, \ldots, N\}$ may take in any generic year $t \in \{1, \ldots, T\}$. Broadly following Dorn and Egger (2011), let us refer by s_{ijt} to a realization of a specific treatment state of pair ij in year t of the 32-nomial variable S_t, containing all possible states. Hence, at any time t there is a sequence of T_t 32-nomial random variables, $\underline{S}_t = (S_1, \ldots, S_t)$, and a sequence of specific realizations of $\underline{s}_{ijt} = (s_{ij1}, \ldots, s_{ijt})$ with $\underline{s}_{ijt}$. Clearly, as time marches on, $\underline{s}_{ijt}$ is drawn from an enormously large set of 32^t possible treatment transitions. Each sequence $\underline{s}_{ijt}$ is associated with potential outcomes y_{ijt}. For a given realization $\underline{s}_{ijt}$ of sequence $\underline{S}_t$, let us refer to the potential outcome in levels as $y_{ijt}^{\underline{s}}$.

Since country-pairs are only observed in one treatment state at a time, we have to impute (estimate) unobserved counterfactual states. Let us denote those counterfactual states at time t by $\tilde{s}_{ijt}$. Using the same notation, let us refer to the counterfactual realization of the random sequence $\underline{S}$ by $\underline{\tilde{s}}_{ijt}$, and to the corresponding counterfactual outcomes by $y_{ijt}^{\underline{\tilde{s}}}$.

Suppose there is a $K \times 1$ random vector of covariates $\mathbf{X}$ with realization $\mathbf{x}_{ijt}$ in t for country-pair ij. We assume that country-pair ij chooses sequence $\underline{S}_t = \underline{s}_{ijt}$ versus $\underline{S}_t = \underline{\tilde{s}}_{ijt}$, depending on $\mathbf{x}_{ijt}$.[22] Following Dorn and Egger (2011), χ may denote the support of values $\mathbf{x}_{ijt}$ the random vector $\mathbf{X}$ can take on.

With multiple treatment effects (i.e. S_t having more than two elements), there are several types of treatment effects one might consider (see Lechner, 2001). First of all, there is a treatment effect on the treated (ATT), which is defined as $\theta_t^{\underline{s},\underline{\tilde{s}}}(\mathbf{X} = \mathbf{x}_{ijt} \in \chi) = E(y_{ijt}^{\underline{s}} - y_{ijt}^{\underline{\tilde{s}}})$ for all $\mathbf{x}_{ijt} \in \chi$ and all $ij \in \{1, \ldots, N^{\underline{s}}\}$ or, in short, $\theta_{ijt}^{\underline{s},\underline{\tilde{s}}}$.

[22] Implicitly, this assumes that $\mathbf{x}_{ijt}$ contains relevant information about $\mathbf{x}_{ijs}$ with $s < t$. This can be ensured by including lagged values of the covariates in $\mathbf{x}_{ijt}$, and/or by conditioning on treatment states prior to t.

Then, there are two concepts of average treatment effects. One is a weighted average of $\theta_{ijt}^{\underline{s},\underline{\tilde{s}}}$ and $\theta_{ijt}^{\underline{\tilde{s}},\underline{s}}$. Following Lechner (2001), we may refer to it by $\alpha_{ijt}^{\underline{s},\underline{\tilde{s}}}$. A second one is a weighted average not only of $\theta_{ijt}^{\underline{s},\underline{\tilde{s}}}$, and $\theta_{ijt}^{\underline{\tilde{s}},\underline{s}}$ but of all $\theta_{ijt}^{\underline{s},\underline{\tilde{s}}'}$ and $\theta_{ijt}^{\underline{\tilde{s}}',\underline{s}}$, which we might call $\gamma_{ijt}^{\underline{s},\underline{\tilde{s}}}$. Notice that, with a 32-nomial treatment problem, there are 32-1 variants of $\tilde{s}'$ (i.e. the states in year t given a common history) and, hence, a gigantic amount of variants of $\underline{\tilde{s}}'$ (i.e. forms of the whole history up to t).

Clearly, with 32 possible treatment states and 46 years, the amount of possible treatment effect estimates is virtually indigestible. Therefore, we decided to focus mainly on ATTs of levels for the year 2005, disregarding country-pairs' histories, and, alternatively, on ATTs with a common history within a well-defined time window. For ATTs in 2005, we use treated and control units only for that year. For ATTs in any year t, we use an 11-year, centred, consecutive time window around the date of switching into treatment. Then, we compare treated units which changed treatment in the sixth consecutive year with ones that did not switch their status within the same time window (i.e. nor before nor after the treated switched).

A.2. A BRIEF SUMMARY OF ASSUMPTIONS

For a causal interpretation of the estimated treatment effects, let us briefly and largely informally state the necessary assumptions (see Dorn and Egger, 2011, for a formal statement of those assumptions in a different context).

- **Assumption 1: Conditional independence**. We assume that we may restore independence of units of observation by conditioning on a vector of observable variables $\mathbf{X} = \mathbf{x}_{ijt}$ and, eventually, on the treatment state prior to switching (see Lechner, 2001; Lee, 2005).
- **Assumption 2: Exogeneity of the initial state and stationary one-period transition probabilities**. By this assumption, one can simply condition on the initial treatment status (i.e. the one prior to t) when estimating treatment effects in t (see Lechner, 2004, for a version of this assumption). Furthermore, the marginal probability of pair ij to be observed with treatment status $S_t = s_{ijt}$ is a first-order stationary Markov-chain, once we condition on state $S_s = s_{ijs}$ with $s < t$ and $\mathbf{X} = \mathbf{x}_{ijt}$.
- **Assumption 3: Balancing score**. This assumption ensures that we can use the scalar-valued function $p\ (\mathbf{X} = \mathbf{x}_{ijt}\ )$ instead of all elements in $\mathbf{X} = \mathbf{x}_{ijt}$ separately to solve the selection problem.
- **Assumption 4: Common support**. We have to ensure that sufficiently comparable control units – i.e. ones with sufficiently similar characteristics $\mathbf{X} = \mathbf{x}_{ijt}$ and balancing score $p^{\underline{s}}\ (\mathbf{X} = \mathbf{x}_{ijt})$ – are used to impute the unobserved counterfactual treatment state of treated units and vice versa.

Based on Assumptions 1–4, all treatment effects of interest are pair-wise identified conditional on the balancing score $p^{s}\ (\mathbf{X} = \mathbf{x}_{ijt}\ )$ and the initial treatment state (in

$s < t$). Provided this, we may employ matching on p^s ($\mathbf{X} = \mathbf{x}_{ijt}$) to resolve the problem of self-selection of country-pairs into the 32 treatment states.

A.3. PROPENSITY SCORE ESTIMATION AND MATCHING ALGORITHM

A.3.1. Propensity score estimation algorithm

For the treatment problem with more than two treatment states at hand, $p^s_{ijt} = p^s(\mathbf{X} = \mathbf{x}_{ijt})$ has to be estimated from multinomial choice models. Obvious candidates are a 32-nomial logit model based on 32 mutually exclusive treatment states[23] or a 5-variate probit model with potentially correlated treatment indicators. We chose to follow the latter approach. Moreover, with 46 years covered, p^s ($\mathbf{X} = \mathbf{x}_{ijt}$) can be estimated from a pooled model or from annual or clusters-of-years choice models. Due to better convergence properties, we estimated a pooled time-series cross-section multivariate probit model. Obviously, there is a strong time pattern in the choices. We accounted for the latter by basing the index (or latent process) underlying the choice model on a specification of $\mathbf{X} = \mathbf{x}_{ijt}$ which includes time averages per country-pair of all the covariates (following the so-called Mundlak–Chamberlain–Wooldridge device)[24] plus a common time trend. Thus, we estimate *seemingly unrelated regression* non-linear probability models which account for the interdependence of the choices of the five modes not only by means of explanatory variables but also in terms of unobservables captured by the disturbances. For the 5-variate probit model based on 5 correlated equations, let us specify the latent process for the latent variable $h^*_{\ell,ijt}$ of equation (i.e., PEIA mode) ℓ, country – pair ij, and time t as a linear function of the $1 \times K$ vector $\mathbf{x}_{ijt}$.[25] Let us denote the $K \times 1$ vector of unknown parameters on $\mathbf{x}_{ijt}$ in equation ℓ by $\boldsymbol{\beta}_\ell$. Finally, let us denote the stochastic term by $\varepsilon_{\ell,ijt}.h^*_{\ell,ijt}$ generates an indicator variable for the ℓth PEIA mode through the function $h_{\ell,ijt} = 1[h^*_{\ell,ijt} > 0]$. Stacking the model for all country-pairs ij and years t and writing vectors in bold-face letters, we may state the model for mode ℓ as

$$\mathbf{h}^*_\ell = \mathbf{x}\boldsymbol{\beta}_\ell + \varepsilon_\ell, \mathbf{h}_\ell = 1[\mathbf{h}^*_\ell > 0], \quad \ell = 1, \ldots, 5$$

The error terms $\varepsilon_{\ell,ijt}$, $\ell = 1, \ldots, 5$ are assumed to be distributed as multivariate normal with zero mean and a variance-covariance matrix $\mathbf{V}$. Specific to this model is that it allows the off-diagonal elements of $\mathbf{V}$ to be non-zero. The parameters $\boldsymbol{\beta}_\ell$ and the error correlations among the five equations along with the 32

[23] Recall that only 26 of the 32 options in Table 2 are actually taken, so that one would actually estimate a 26-nomial logit model.

[24] Hence, the model includes fixed country-pair effects which are parameterized as a linear function of time-averaged observables (see Wooldridge, 2002).

[25] In a reduced form of the 5-variate latent process, we may suppress an equation index with $\mathbf{x}_{ijt}$ and allow the latent process in each equation $\hbar$ to depend on the same observables without loss of generality.

$\sum_{t=1}^{T} N_t \times 1$ vectors $\mathbf{p}^s$ with elements $p_{ijt}^s = p^s(\mathbf{X} = \mathbf{x}_{ijt})$ can be estimated by Monte Carlo integration of the maximum likelihood function and a 5-variate normal distribution, respectively.[26] For this, we followed Johnson *et al.* (2000) and Craig (2008) in employing the Gauss–Legendre method. Prediction of the propensities of participating in a particular type of PEIA and, hence, propensities for all 32 (or 26 actually taken) treatments as in Table 2 can be retrieved from Monte Carlo integration as well.

A.3.2. Matching algorithm

With 32 vectors at hand, we may impute 31 possible counterfactual states for each treatment state a country-pair ij actually has in any year t, provided that there are enough treated and control units available and that the scalar-valued balancing score indeed balances the underlying observables. In order to determine comparable units in $\mathbf{p}^s$-space, we have to impose a rule determining which units are comparable and which are not. Commonly adopted procedures are M-nearest neighbor matching, where the researcher *ex ante* determines the integer M but not the minimum tolerable degree of similarity, and kernel matching, where the researcher *ex ante* determines the minimum tolerable degree of similarity between treated and control units but not the number of compared (matched) controls. So-called radius matching may be viewed as a special case of kernel matching (with a fixed bandwidth and a uniform kernel). The latter is what we employ, with a chosen radius of 0.01.[27]

In order to reduce the curse of dimensionality given the large possible treatment status space and the potentially enormous demands on computing in order to ensure balancing in an approach as in Dorn and Egger (2011), we implement estimation of the average treatment effects of the treated (ATTs) as follows. First, radius matching in a subsample of units with treatment status s_{ijt} or $\tilde{s}_{ijt}$ and, hence, a propensity score subvector for those two types of units only, say $\mathbf{p}_{\tilde{s}}^s$, obtains kernel weights for the treated (s) and the control observations ($\tilde{s}$).[28] These weights reflect which units are used for comparison at all and which are not, and they reflect the number of times the same control units are used for comparison (e.g. multiple use of control units leads to higher

[26] Note that N_t in $\sum_{t=1}^{T} N_t \times 1$ acknowledges the fact that not all country-pairs are covered in all of the 46 years, so that the panel dataset is unbalanced. One reason for unbalancedness is that not all countries of 2005 existed in all years covered. Another reason is that control variables are not available for all years and country-pairs.

[27] As usual, the form of the kernel tends to be less critical than the bandwidth. Ideally, one would select the optimal bandwidth by a rule-of-thumb, a grid search, cross validation, or some other loss-function-based approach to trade off precision and efficiency of the estimates. However, the dimensionality of the problem requires us to cut short on this issue and choose a small-enough radius (bandwidth) heuristically.

[28] Recall that we adopt radius matching with a radius of 0.01. For estimating, for example, the average treatment effect on the treated of a GTA relative to a counterfactual with no PEIA at all, we compare country-pairs with a GTA and propensity score sub-vector $\mathbf{p}_s^s$ with units that actually got no-PEIA treatment and had propensities for getting treatment of a GTA, $\mathbf{p}_{\tilde{s}}^s$, which fulfil $|\mathbf{p}_s^s - \mathbf{p}_{\tilde{s}}^s| \leq 0.01\iota$, where ι is a vector of ones of proper size.

standard errors through the weighting). For ATTs in the year 2005, $\theta^{s,\tilde{s}}_{2005}$,[29] we then run weighted least squares. For this, we define an $N_{2005} \times (K+1)$ matrix $\mathbf{Z}_{2005}$. The first column of $\mathbf{Z}_{2005}$ contains an indicator variable which is unity if pair ij has actual treatment status $s_{ij,2005}$ and zero else, and the other columns of $\mathbf{Z}_{2005}$ contain the elements of $\mathbf{x}_{ij,2005}$. Then, we define an $N_{2005} \times N_{2005}$ matrix containing the weights from radius matching, $\mathbf{W}_{2005}$. Finally, we obtain the ATT of treatment s relative to $\tilde{s}$, $\theta^{s,\tilde{s}}_{2005}$, as the first element of the $(K+1) \times 1$ vector

$$\hat{\beta} = (\mathbf{Z}'_{2005}\mathbf{W}_{2005}\mathbf{Z}_{2005})^{-1}\mathbf{Z}'_{2005}\mathbf{W}_{2005}\mathbf{y}^{s}_{2005},$$

where $\mathbf{y}^{s}_{2005}$ is an $N_{2005} \times 1$ outcome vector. Notice that conditioning on $\mathbf{x}_{ij,2005}$ in the estimation of ATT reduces the potential bias from a possible lack of balancing in $\mathbf{p}^{s}$ (see Blundell and Costa Dias, 2009).[30]

With panel data and switching into treatment of country-pair ij in an arbitrary year t, we use the vector $\mathbf{p}^{s}_{t}$ for that pair to select control units in t with identical treatment status $\tilde{s}$ prior to treatment switching into status s in all years within the centred window around the treatment change. We match those control units in the year of the treatment change and then estimate weighted least squares regressions using the radius matching weights for all matched pairs with the same initial treatment state. We do so for the most important (in terms of numbers of switching cases from Table 2) treatment switches. The corresponding treatment effects may be labelled $\theta^{s,\tilde{s}}_{t}$, since they condition at least on part of the history of integration in the time neighbourhood of treatment switching.[31]

APPENDIX B

GENERAL EQUILIBRIUM EFFECTS

In multi-country models of goods trade, services trade or foreign direct investment, third-country effects of fundamentals arise naturally. Baier and Bergstrand (2009) provide a simple, linearized form of the effects of trade friction and trade facilitation variables in new trade theory models (for examples see Eaton and Kortum, 2002;

[29] Notice that we speak now of $\theta^{s,\tilde{s}}_{2005}$ rather than $\theta^{s,\tilde{s}}_{2005}$ to emphasize that we *do not* condition the integration history explicitly here.

[30] For estimates of long-run responses of outcome, we do not specify a time frame within which the response has to take place. We can identify long-run responses which do not condition on the treatment effect history from the cross-sectional variation in the data. There, the treatment status could principally be acquired and the treatment effect be realized even before the sample period.

[31] Short-run responses have to take place within a certain time frame after switching into a new treatment status. With short-run responses, we condition on the treatment status history in the time neighbourhood of switching. The latter implies that we can only analyse a subset of treatment effects, since case numbers of changes are too small for most treatments (see Table 2).

Anderson and van Wincoop, 2003). This paper focuses on the estimation of ATTs of PEIAs (dubbed $\theta^{\underline{s},\underline{\tilde{s}}}$ in Appendix A). These treatment effects are asymmetric across alternatives in the sense that $\theta^{\underline{s},\underline{\tilde{s}}}$ does not have to equal $-\theta^{\underline{\tilde{s}},\underline{s}}$. An average treatment effect (ATE) which is symmetric in that sense was introduced as $\gamma^{\underline{s},\underline{\tilde{s}}}$ in Appendix A. In cross-section or panel data models, parameters on PEIA modes are commonly interpreted as estimates of $\gamma^{\underline{s},\underline{\tilde{s}}}$. Let us briefly outline how the *direct* ATE $\gamma^{\underline{s},\underline{\tilde{s}}}$ (and, by a similar token, the direct ATT $\theta^{\underline{s},\underline{\tilde{s}}}$) relates to the *total* ATE (or ATT), which accounts for general equilibrium effects. For this, it is useful to assume cross-sectional data so that the time index can be dropped. Suppose we are interested in the effect of some generic PEIA indicator T_{ij} for, say, GTA membership or STA membership. Furthermore, let ϕ_i denote country i's share of GDP in world GDP. Then, Baier and Bergstrand (2009) suggest that OLS regressions could be based on log-transformed normalized bilateral trade flows and transformed trade cost variables in order to render coefficients of trade costs consistent in spite of the presence of general equilibrium effects. The transformed counterpart to T_{ij} would be:

$$\tilde{T}_{ij} = T_{ij} - \sum_{i=1}^{J} \phi_i T_{ij} - \sum_{j=1}^{J} \phi_j T_{ij} + \sum_{i=1}^{J} \sum_{j=1}^{J} \phi_i \phi_j T_{ij}$$

where J denotes the number of countries in the world economy. The parameter on $\tilde{T}_{ij}$ in an OLS regression with (GDP-normalized) log bilateral trade flows as the dependent variable would then be a consistent estimate of the treatment effect $\gamma^{\underline{s},\underline{\tilde{s}}}$ as a *direct effect*. In this paper, we are interested in estimating *direct effects* of switching from $T_{ij} = 0$ to $T_{ij} = 1$ for country pair ij in the sense of $\theta^{\underline{s},\underline{\tilde{s}}}$(or $\gamma^{\underline{s},\underline{\tilde{s}}}$). Clearly, the total effect cum general equilibrium repercussions of such a change for trade from country i to j amounts to the respective treatment effect times

$$\tilde{T}_{ij}(T_{ij} = 1) - \tilde{T}_{ij}(T_{ij} = 0) = (1 - 2\phi_i - 2\phi_j + 2\phi_i\phi_j)$$

Notice that the presence of $(-2\phi_i - 2\phi_j + 2\phi_i\phi_j)$ in that formula indicates that PEIA treatments are symmetric so that i and j arise as an exporter and as an importer each. A respective positive total effect of treatment T_{ij} will be smaller than the direct effect and depend on ϕ_i and ϕ_j. However, on average GDP shares ϕ_i are small so that $\theta^{\underline{s},\underline{\tilde{s}}}$ and $\gamma^{\underline{s},\underline{\tilde{s}}}$ will approximately measure direct as well as total effects unless very large economies are at stake.

REFERENCES

Anderson, J.E. and E. van Wincoop (2003). 'Gravity with gravitas: a solution to the border puzzle', *American Economic Review*, 93(1), 170–92.

Arcand, J.-L., M. Olarreaga and L. Zoratto (2010). 'Weak governments and trade agreements', unpublished manuscript, University of Geneva.

Baier, S.L. and J.H. Bergstrand (2004). 'Economic determinants of free trade agreements', *Journal of International Economics*, 64, 29–63.

—— (2007). 'Do free trade agreements actually increase members' international trade?' *Journal of International Economics*, 71(1), 72–95.
—— (2009). 'Estimating the effects of free trade agreements on international trade flows using matching econometrics', *Journal of International Economics*, 77(1), 63–76.
Baldwin, R. (1993). 'A domino theory of regionalism', NBER Working Paper No. 4465.
—— (2011). '21st century regionalism: Filling the gap between 21st century trade and 20th century trade rules', CEPR Policy Insight No. 56, Centre for Economic Policy Research, London.
Barro, R. and S. Tenreyro (2007). 'Economic effects of currency unions', *Economic Inquiry*, 45 (1), 1–23.
Bergstrand, J.H. and P. Egger (2007). 'A knowledge-and-physical-capital model of international trade flows, foreign direct investment, and multinational enterprises', *Journal of International Economics*, 73(2), 278–308.
—— (2011). The determinants of BITs, unpublished manuscript.
Bergstrand, J.H., P. Egger and M. Larch (2010). 'The timing of preferential trade agreement membership', unpublished manuscript.
Blonigen, B. and R.B. Davies (2004). 'The effects of bilateral tax treaties on U.S. FDI activity', *International Tax and Public Finance*, 11(5), 601–22.
Blundell, R. and M. Costa Dias (2009). 'Alternative approaches to evaluation in empirical microeconomics', *Journal of Human Resources*, 44(3), 565–640.
Conconi, P., A. Sapir and M. Zanardi (2012). 'The internationalization process of firms: from exports to FDI', mimeo, Université Libre de Bruxelles.
Craig, P. (2008). 'A new reconstruction of multivariate normal orthant probabilities', *Journal of the Royal Statistical Society Series B*, 70(1), 227–43.
Dorn, S. and P. Egger (2011). 'On the distribution of exchange rate regime treatment effects on international trade', CEPR Discussion Paper No. 8654.
Eaton, J. and S. Kortum (2002). 'Technology, geography, and trade', *Econometrica*, 70(5), 1741–79.
Egger, P. (2008). 'De facto exchange rate arrangement tightness and bilateral trade flows', *Economics Letters*, 99(2), 228–32.
Egger, H., P. Egger and D. Greenaway (2008). 'Trade liberalization with multinational firms: effects on welfare and intra-industry trade', *Journal of International Economics*, 74(2), 278–98.
Egger, P. and R. Lanz (2008). 'The determinants of GATS commitment coverage', *The World Economy*, 31(12), 1666–94.
Egger, P. and M. Larch (2008). 'Interdependent preferential trade agreement memberships: an empirical analysis', *Journal of International Economics*, 76(2), 384–99.
Egger, P., M. Larch and M. Pfaffermayr (2007a). 'On the welfare effects of trade and investment liberalization', *European Economic Review*, 51, 669–94.
—— (2007b). 'Bilateral versus multilateral trade and investment liberalization', *The World Economy*, 30(4), 567–96.
Egger, P., M. Larch, M. Pfaffermayr and H. Winner (2006). 'The impact of endogenous tax treaties on foreign direct investment: theory and empirical evidence', *Canadian Journal of Economics*, 39(3), 901–31.
Egger, P., M. Larch, K. Staub and R. Winkelmann (2011). 'The trade effects of endogenous preferential trade agreements', *American Economic Journal: Economic Policy*, 3(3), 113–43.
Egger, P. and V. Merlo (2012). 'BITs bite: an anatomy of the impact of bilateral investment treaties on multinational firms', *Scandinavian Journal of Economics*, 114(4), 1240–66.
Egger, P. and D. Nelson (2010). 'How bad is antidumping? Evidence from panel data', *Review of Economics and Statistics*, 93(4), 1374–90.
Egger, P. and M. Pfaffermayr (2004). 'The impact of bilateral investment treaties on foreign direct investment', *Journal of Comparative Economics*, 32(4), 788–804.
Ekholm, K., R. Forslid and J.R. Markusen (2007). 'Export-platform foreign direct investment', *Journal of the European Economic Association*, 5(4), 776–95.
Francois, J. and B. Hoekman (2010). 'Services trade and policy', *Journal of Economic Literature*, 48 (3), 642–92.
Francois, J., O. Pindyuk and J. Wörz (2009). Trends in international trade and FDI in services: a global database of services trade, IIDE Discussion Papers No. 20090802.

Glick, R. and A.K. Rose (2002). 'Does a currency union affect trade? The time-series evidence', *European Economic Review*, 46(6), 1125–51.
Goldberg, L. and C. Kolstad (1995). 'Foreign direct investment, exchange rate variability and demand uncertainty', *International Economic Review*, 36(4), 855–73.
Head, K. and J. Ries (2004). 'Exporting and FDI as alternative strategies', *Oxford Review of Economic Policy*, 20, 409–23.
Helpman, E. (1984). 'A simple theory of international trade with multinational corporations', *Journal of Political Economy*, 92(3), 451–71.
Helpman, E. and P. Krugman(1985). *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*, MIT Press, Cambridge, MA.
Horn, H., P.C. Mavroidis and A. Sapir (2010). 'Beyond the WTO? An anatomy of EU and US preferential trade agreements', *World Economy*, 33(11), 1565–88.
Johnson, N.L., S. Kotz and N. Balakrishnan (2000). *Continuous Multivariate Distributions, Volume 1: Models and Applications*, 2nd edn, Wiley Series in Probability and Statistics.
Lechner, M. (2001), 'Identification and estimation of causal effects of multiple treatments under the conditional independence assumption', in M. Lechner and F. Pfeiffer (eds.), *Econometric Evaluation of Labour Market Policies*, Physica Verlag, Heidelberg, 43–58.
—— (2004). 'Sequential matching estimation of dynamic causal models', IZA Discussion Paper 1042.
Lee, M.-J. (2005). *Micro-Econometrics for Policy, Program, and Treatment Effects*, Oxford University Press, New York.
Magee, C.S. (2003). 'Endogenous preferential trade agreements: an empirical analysis', *Contributions to Economic Analysis & Policy* 2, Article 15.
Markusen, J.R. (1984). 'Multinationals, multi-plant economies, and the gains from trade', *Journal of International Economics*, 16(3–4), 205–26.
—— (1995). 'The boundaries of multinational enterprises and the theory of international trade', *Journal of Economic Perspectives*, 9(2), 169–89.
—— (2002). *Multinational Firms and the Theory of International Trade*, MIT Press, Cambridge, MA.
Markusen, J.R. and A.J. Venables (1998). 'Multinational firms and the new trade theory', *Journal of International Economics*, 46(2), 183–203.
—— (2000). 'The theory of endowment, intra-industry and multi-national trade', *Journal of International Economics*, 52(2), 209–34.
Marshall, M.G. and K. Jaggers (2002). '*Polity IV project: political regime characteristics and transitions, 1800–2002*', Dataset Users' Manual, Center for International Development and Conflict Management, University of Maryland.
Persson, T. (2001). 'Currency unions and trade: How large is the treatment effect?' *Economic Policy*, 16(33), 433–48.
Reinhart, C. and K. Rogoff (2004). 'The modern history of exchange rate arrangements: a reinterpretation', *Quarterly Journal of Economics*, 119(1), 1–48.
Rose, A.K. (2000). 'One money, one market: the effects of common currencies on trade', *Economic Policy*, 15(30), 7–46.
Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.
—— (2005). 'Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity', *Journal of Applied Econometrics*, 20, 39–54.